INDIAN PHILOSOPHY
A COLLECTION OF READINGS

Metaphysics

Edited by
Roy W. Perrett

ROUTLEDGE

Indian Philosophy

A Collection of Readings

Series Editor

Roy W. Perrett

Massey University

Series Contents

Metaphysics

Edited with introductions by

Roy W. Perrett

Massey University

Routledge
Taylor & Francis Group
New York London

Routledge
Taylor & Francis Group
270 Madison Avenue
New York, NY 10016

Routledge
Taylor & Francis Group
2 Park Square
Milton Park, Abingdon
Oxon OX14 4RN

Garland is an imprint of the Taylor & Francis Group

The Library of Congress Cataloging-in-Publication Data

Indian Philosophy / edited with introductions by Roy W. Perrett.
p.cm.— (Indian Philosophy ; 3)
Includes bibliographical references.
ISBN 0-8153-3608-X (alk. paper)
1. Philosophy, Indic. 2. Metaphysics. I. Perrett, Roy W. II. Series.

B130.I63 2000 vol. 3
181'.4s—dc21
[110'.954] 00-0062310

5 volume set ISBN:	**0-8153-33670-1**
Volume 1: Epistemology	0-8153-3609-8
Volume 2: Logic and Philosophy of Language	0-8153-3610-1
Volume 3: Metaphysics	0-8153-3608-X
Volume 4: Philosophy of Religion	0-8153-3611-X
Volume 5: Theory of Value	0-8153-3612-8

Contents

Series Preface

No anthologist succeeds in including everyone's favorites, so a few words about the principles of selection seem appropriate. Firstly, as with other volumes in this Garland series, priority has been given to journal articles, rather than book chapters. However, some essential book chapters have been included, and the introductions to each volume include references to significant books. Readers in search of further bibliographical assistance should consult what is now the standard source: Karl H. Potter, *Bibliography of Indian Philosophies*, 3rd rev. ed. (Delhi: Motilal Banarsidass, 1995), and the on-line updates to it available at the "Indian Philosophy Bibliography" (http://faculty.washington.edu/kpotter/). Secondly, the emphasis throughout is on *philosophical* studies of Indian philosophy. Consequently, much excellent historical and philological work has been omitted. Thirdly, the desire to make Indian philosophy accessible to interested Western philosophers has meant not only that all the selections are in English, but also that most of them use a minimal amount of unglossed Sanskrit terminology. This restriction has prevented the inclusion of more work by Indian authors.

Finally, I would like to acknowledge gratefully all the good advice and generous assistance I have received from Chakravarthi Ram-Prasad, Mark Siderits, and (most especially) Stephen Phillips. The blame for any shortcomings that may remain rests, of course, solely upon my own shoulders.

This page intentionally left blank

Series Introduction

The five volumes of this series collect together some of the most significant modern contributions to the study of Indian philosophy. Indian philosophy is one of the great philosophical traditions of the world. Unfortunately, however, its philosophical riches are not always as readily accessible to Western philosophers as might be desired. The selections in these volumes help to redress this situation by giving readers easy access to some of the best philosophical work in the area, including material that is often difficult to locate.

Collectively the selections in these volumes explore many of the important commonalities and differences between the Indian and Western philosophical traditions. These similarities and differences are philosophically pregnant. There is enough in common between Indian and Western philosophy to suggest that the philosophers in both traditions are often engaged with similar problems and hence should be able to communicate with each other. However, there are also sufficient differences between the traditions to suggest that they may have some novel perspectives to offer each other.

In choosing the selections for this series priority has been given to journal articles, rather than book chapters. However, some essential book chapters have been included, and the introductions to each volume include references to significant books. The emphasis throughout is on philosophical studies of Indian philosophy. The desire to make Indian philosophy accessible to interested Western philosophers has meant not only that all the selections are in English, but also that most of them use a minimal amount of unglossed Sanskrit terminology.

Volume 1: Epistemology is concerned with the nature and scope of Indian pramāṇa theory, i.e. that part of Indian philosophy concerned with the nature and sources of knowledge. Indian philosophers developed a causal theory of knowledge and acknowledged the existence of a number of valid ways of knowing, including perception, inference and testimony. The Indian pramāṇa theorists thus discussed many issues that have also occupied Western epistemologists, often offering importantly different perspectives on these

matters. They also sometimes addressed various interesting questions about knowledge that are unfamiliar to Western epistemologists.

The selections in this volume discuss Indian treatments of epistemological topics like the means of knowledge, realism and anti-realism, truth, knowledge of knowledge, illusion and perceptual error, knowability, testimony, scepticism and doubt.

Volume 2: Logic and Philosophy of Language is concerned with those parts of Indian pramāṇa theory that Western philosophers would count as logic and philosophy of language. Indian philosophers and linguists were much concerned with philosophical issues to do with language, especially with theories of meaning, while the Indian logicians developed both a formalised canonical inference schema and a theory of fallacies. The logic of the standard Indian inferential model is deductive, but the premises are arrived at inductively. The later Navya-Nyāya logicians went on to develop too a powerful technical language, an intentional logic of cognitions, which became the language of all serious discourse in India.

The selections in this volume discuss Indian treatments of topics in logic and the philosophy of language like the nature of inference, negation, necessity, counterfactual reasoning, many-valued logics, theory of meaning, reference and existence, compositionality and contextualism, the sense-reference distinction, and the nature of the signification relation.

Volume 3: Metaphysics is concerned with the complement to pramāṇa theory, i.e. prameya theory. Whereas the pramāṇas are the means of knowledge, the prameyas are the knowables, cognisable entities that constitute the world. With respect to the number and kinds of such entities, there was a very wide variety of opinion among classical Indian philosophers — including variants of monism, dualism and pluralism about both entities and kinds. Many metaphysical topics were debated, but two of the most important were causation and the nature of the self. The competing theories offered about these two issues also raised other questions about the metaphysics of wholes and parts, substances and properties, and universals and particulars.

The selections in this volume discuss Indian treatments of topics in metaphysics like ontology, constructionalism, universals, negative facts, mereology, causation, relations, freedom and determinism, and theories of the self.

Volume 4: Philosophy of Religion is concerned with something that can be described as "Indian philosophy of religion," i.e. "philosophy of Indian religions." Contrary to popular Western belief, classical Indian philosophy was not indistinguishable from Indian religion — as even a cursory glance at the first three volumes of this series will demonstrate. Religious concerns, though, did motivate the work of many Indian philosophers. However, important differences between the major Western religions and the major Indian religions (Hinduism, Buddhism, Jainism) mean that the shape of Indian philosophy of religion is often significantly different from that of Western philosophy of religion.

The selections in this volume discuss Indian treatments of topics in the philosophy of religion like the problem of evil, God, theological monism and dualism, atheism, the concept of a perfect being, reason and revelation, rebirth and karma, religious language, religion and politics, ritual and mantra, and the religious determinants of metaphysics.

Volume 5: Theory of Value is concerned with Indian discussions in the areas of ethics, politics and aesthetics. The Indian philosophers had a good deal to say about the theory of value as they vigorously discussed topics like the ends of life and the relation of virtuous action to those ends. A traditional Hindu classification recognises four classes of values: dharma (morality, virtue), artha (wealth, power), kāma (pleasure), and mokṣa (liberation). Mokṣa is usually held to be the highest value and is extensively discussed in the paradigm Indian philosophical texts. Indian political and legal theory is concerned with the values of artha and dharma. Aesthetic pleasure is one of the subject matters of a developed body of writing on aesthetic theory. Rasa ("flavor"), the special feeling or enjoyment that pervades an artwork or is aroused in its contemplator, is commonly seen as detached from the aims and concerns of ordinary life, with some even suggesting that it provides a foretaste of the bliss of mokṣa.

The selections in this volume discuss Indian treatments of topics in the theory of value like the proper ends of life, the relation of dharma and mokṣa, liberation and pleasure, the sources of our knowledge of right and wrong, the ethics of non-violence, the status of the supra-moral, egoism and altruism, the theory of rasa, aesthetic experience and catharsis.

This page intentionally left blank

Volume Introduction

The complement to *pramāṇa* theory is *prameya* theory. Whereas the *pramāṇas* are the means of knowledge, the *prameyas* are the knowables, cognizable entities which constitute the world. With respect to the number and kinds of such entities, there was a very wide variety of opinion among classical Indian philosophers. Moreover, since according to most Indian systems knowledge of reality is at least a necessary condition for liberation, these metaphysical disputes were taken to be of practical as well as theoretical import.

Ontology is an attempt to answer the question "What is there?". But to answer that question we need to distinguish between the two separate, though intertwined, questions: "How many entities are there?" and "How many *kinds* of entities are there?". In both cases the metaphysically interesting answers are: "One", "Two", and "Many" (i.e. monism, dualism, and pluralism). Note, however, that a position about the number of kinds of entities that exist does not in itself entail any particular position about the number of entities that exist. Nor does a dualism or pluralism about the number of entities entail any particular position about the number of kinds of entities.

There was, correspondingly, quite a variety of Indian responses to the question "What is there?", including variants of monism, dualism and pluralism about both entities and kinds. Advaita, for instance, holds that there is numerically only one entity (*ātman/Brahman*) and that all plurality is illusory. Viśiṣṭādvaita qualifies this monism and maintains that while there exists only one ontologically independent substance (God), there also exist other dependent entities (souls and material objects). Sautrāntika Buddhism, in contrast, holds that there are numerically many entities, but only one *kind* of thing: momentary particulars (*svalakṣaṇas*). Yogācāra Buddhism and Cārvāka also agree that there is only one kind of thing, but disagree about whether it is mental or material. Sāṃkhya-Yoga, on the other hand, asserts that reality consists of just two kinds of things but a plurality of entities: many selves (*puruṣas*) and a single evolving primal matter (*prakṛti*). Finally, Nyāya-Vaiśeṣika is pluralist about both entities and kinds: there are many things in the world, though there are only seven basic ontological categories (*padārtha*).

Underpinning these disagreements about the nature and number of reals are also important differences of opinion about the criterion of reality. According to the Buddhists, for instance, to be real is to be causally efficacious. For Advaita, however,

the real is that which is never sublated (*abādha*) by any manner or means. Thus only *Brahman*, free from all limitations of space, time and individuality, is truly real.

In their systematic development of these differing ontologies Indian philosophers mostly utilized a common methodology. Since even the most revisionary Indian metaphysicians usually took seriously the defeasible deliverances of common sense, they often conceived of their task as the construction of a philosophical system which permitted all objects recognized by common sense to be reduced to logical constructions out of the favored primitive entities of the system. This general methodological stance is equally true, for example, of the Buddhist logicians (who favored an event ontology of momentary particulars), of Nyāya-Vaiśeṣika (who favored a rich ontology of univerals, qualities and substances), and of late Advaita (who favored an entity monism). Ontological disputes were less about the reductionist project (Madhyamaka Buddhism is a notable exception), than about the details of the attempted reductions: e.g. do we really need to posit wholes as well as parts, substances as well as properties, universals as well as particulars, selves as well as mental states, and so on?

Many metaphysical topics were debated in classical India, but two of the most important were causation and the nature of the self. The Indian philosophers extensively discussed a number of issues relating to causation, including the nature of the causal relation, the definitions of cause and effect, and classifications of kinds of causes (Potter 1963, Smart 1964, Bhartiya 1976). Typically they stressed the importance of the material cause, rather than (as in Western philosophy) the efficient cause.

The Indian theories of causation are traditionally classified by reference to the question of whether the effect is a mode of the cause. According to this taxonomy there are two principal theories of causation. One is the identity theory (*satkāryavāda)* which holds that the effect is identical with the cause, a manifestation of what is potential in the cause. This is the Sāṃkhya-Yoga view, though that school's particular version of it is sometimes called transformation theory (*pariṇāmavāda*). Advaita Vedånta holds an appearance theory (*vivartavāda*) which is often considered a variant of the identity theory. According to the appearance theory effects are mere appearances of the underlying reality, *Brahman*. Since only *Brahman* truly exists, this theory is also sometimes called *satkāraṇavāda* (the theory that the cause is real but the effect is not).

The other principal theory of causation is the non-identity theory (*asatkāryavāda*) which denies that the effect pre-exists in its cause and claims instead that the effect is an altogether new entity. Both Nyāya-Vaiśeṣika and the Buddhists are usually classified as non-identity theorists, but they differ on many important details. One of these is whether the cause continues to exist after the appearance of the effect: Nyāya-Vaiśeṣika claims it does, the Buddhists mostly claim it does not.

Indian theories of the self also traditionally divide into two broad classes: those who explain diachronic personal identity by reference to an enduring substantial self (*ātmavādins*) and those who deny the existence of such a self, taking instead a "modal" view of reality (*anātmavādins*). The orthodox Hindu philosophers and the Jainas all take the former view. Although they disagree on the nature and number of such selves, they are all non-reductionists of some sort about personal identity. Most Indian Buddhist philosophers (including the Theravādins, the Vaibhāṣikas, the Sautrāntikas, the Yogācārins and the Svātantrika-Mādhyamikas) take the latter view and hence are all

plausibly classifiable as reductionists about personal identity.

The selections in this volume open with an article by Matilal which helpfully sets out something of the basic ontological framework of Indian metaphysics. Matilal concentrates on the very different ontologies of Nyāya-Vaiśeṣika, Buddhism and Jainism. He contrasts the anti-substantialism of the Buddhists with the substantialism of Nyāya-Vaiśeṣika, and presents the Jaina conception of reality as an attempt to bring together the opposing viewpoints of the Buddhists (with their emphasis on becoming) and Nyāya-Vaiśeṣika (with their emphasis on being). It is worth noting that choice between the rival metaphysical systems was not made entirely on the basis of the *pramāṇas*. Rather the ontologies of the systems were defended holistically, though they were open to rational criticism (see Mohanty 1992).

Potter's paper on the *Vedāntaparibhāṣā* reveals in a most illuminating way how the conception of philosophy as systematic reconstruction is shared not only by the realist logicians of the Nyāya-Vaiśeṣika school (Potter 1977), but also by the supposedly "mystical" Advaitins. The eighty-four interconnected definitions in Dharmarāja's text are all generated from a small base of primitive notions. The terms so defined are then used to constitute a rigorous philosophical system, in a fashion somewhat analogous to the constructionalist projects of Carnap and Goodman.

Indian approaches to the problem of universals are addressed in the succeeding articles by Chakrabarti and Siderits. Chakrabarti explains Nyāya-Vaiśeṣika's realist theory of univerals (see also Shastri 1964, Dravid 1972, Potter 1977, Mukhopadhyay 1984, Phillips 1995). According to the Naiyāyikas, universals are both simple and sensible, but there are also various restrictive conditions that prevent the admission of a universal corresponding to every common name. Nyāya metaphysicians accordingly admit universals to their basic ontology, but count as particulars both the inherence relations which bind properties to property-possessors, and various non-repeatable qualities (*guṇas*) such as color, magnitude, weight, shape, etc.

The Nyāya position on universals is thus objectionable both to some fellow realists, who think that qualities and relations should be counted as universals, and to nominalists who think that universals are entirely dispensable. Siderits' paper is concerned with both types of objection. He argues that (*pace* Dravid 1972) there is no incoherence in the notion of the qualities of a particular substance being themselves particulars. He also offers a rational reconstruction of the *apoha* theory of the Buddhist nominalists, who deny that the use of a general term like "cow" ever requires us to posit the existence of a corresponding universal cowness. To say of a particular that it is a cow is just to say that it does not belong to the class of things that are non-cows (see also Shaw 1978).

Nyāya-Vaiśeṣika metaphysicians admit to their categories of the real just six presences: substances, qualities, actions, universals, particularities, and inherence (Potter 1977, Halbfass 1992, Phillips 1995). But they also admit a seventh, and final category: absences (*abhāva*), or negative facts (Matilal 1968, Chakrabarti 1997). Gillon's article draws freely upon both Indian and Western philosophy to present a challenging modern defense of the Nyāya view that absences do indeed exist and that they are known not only by inference, but also by perception.

Any metaphysical reductionism requires a mereology, or theory of the part-

whole relation. In India the two most developed mereological theories were those of Nyāya-Vaiśeṣika and the Buddhists. According to Nyāya-Vaiśeṣika, wholes are never reducible to their parts, though they inhere in them (Bhaduri 1975, Potter 1977). According to the Buddhists, wholes are unreal, mere conceptual constructions, and only parts are real. Kapstein's article reconstructs the Buddhist idealist Vasubandhu's arguments against the atomic theory of the Buddhist realists, arguments which utilize certain implicit mereological principles. Vasubandhu tries to prove idealism by demonstrating that the notion of composite material wholes entails that there are simple, atomic substances, and that atomism is false.

One of the most important relations in any plausible metaphysics is causation. Matilal's article briefly explains the Nyāya-Vaiśeṣika theory of causation, contrasting it with the Ābhidharmika and Sāṃkhya theories. Essential to Nyāya causal theory (and to their metaphysics more generally) is the notion of inherence (*samavāya*), the relation that also connects wholes and parts, and substances and their qualities. The most popular Indian objection to the Nyāya theory of inherence is an infinite regress argument. If two entities *A* and *B* are to be related by the inherence relation *R* which is itself a distinct entity, then it is also necessary that *A* and *R* be related by a different inherence relation *R**, itself a distinct entity. But then, of course, *A* and *R** have to be related by a yet different inherence relation *R***; and so on *ad infinitum*.

The Naiyāyikas reply that there is no regress because there is no other relation to connect inherence to its relatum. Clearly they cannot mean that the relatum and its relation are identical, for then, by the transitivity of identity, *A* would not only be identical with *R* but also with B! Instead Navya-Nyāya appeals to the notion of a self-linking connector (*svarūpasambandha*). The idea here is that while *A* requires the inherence relation *R* to connect it to *B*, *A* can be its own connector to *R*. Potter's paper, although not explicitly about Indian philosophy, is in effect an independent defense of the intelligibility of the notion of such a "concrete connector".

The most Indian famous critique of causality is that of Nāgārjuna, discussed at some length in the piece by Garfield (see also Garfield 1995). In a virtuoso dialectical display Nāgārjuna refutes all the standard types of Indian causal theory: (i) self-causation; (ii) external causation; (iii) both (i.e. self and external causation); and (iv) non-causation. But Nagarjuna does not deny causation *per se*. Rather causality is the Buddha's teaching of dependent origination: everything is interdependent and this is equivalent to the truth of emptiness (*śūnyatā*), that nothing has any self-existence or essence. Liberation is the realization of this emptiness.

Contrary to popular Western belief, the general Indian acceptance of universal causality and karma does not entail fatalism, as Potter explains in the succeeding article. However, there is a widespead Indian commitment to a fairly strong form of determinism. In the course of defending the compatibility of this determinism with a kind of freedom of agency, Potter also develops an original argument for the existence of beginningless agents.

The remainder of the selections are concerned with Indian theories of the self. Central to early Buddhism is the doctrine of "no-self" (*anatta*), which denies the existence of an enduring substantial self (Collins 1982). The Ābhidharmikas developed this doctrine into a detailed reductionist account of persons that has certain parallels

with the theory of personal identity recently developed by Derek Parfit (Parfit 1984). The articles by Duerlinger, Bastow and Siderits explore these parallels in some detail, and reach somewhat different conclusions. Duerlinger (invoking Candrakīrti) argues that reductionism is not the only Buddhist view, while Bastow and Siderits suggest that Buddhist reductionism has resources which can usefully fill lacunae in Parfit's theory.

In contrast to the Buddhists, the Hindu philosophers were all committed to the existence of an enduring substantial self (*ātman*), though there was lively disgreement among them as to the nature and number of such selves. One argument for the *ātman* (to be found in both Mīmāṃsā and Nyāya) is that personal memories of the form "I did this" presuppose the existence of a continuing substantial self. Taber explains the Mīmāṃsā version of the argument, and Chakrabarti defends a modern version of the Nyāya argument (see also Chakrabarti 1999).

The piece by Shukla is interesting for several reasons, including its contrast with Chakrabarti's. Translated from modern Sanskrit, it is a striking example of the vitality of the pandit tradition in India (see also Krishna 1991, Sankaranarayanan 1993). Working entirely within the context of traditional Sanskritic Nyāya scholarship, Shukla argues that considerations of ontological economy (*lāghava*) show that Nyāya should refuse to admit *ātman* as a separate metaphysical category. Everything that the self was postulated to do in the system can be done by the (corporeal) *manas* or "mind" alone. Hence on this matter the Nyāya tradition requires rational revision.

Sāṃkhya-Yoga is the oldest school of Indian philosophy (Larson 1979, Larson and Bhattacharya 1987). It is committed to a sharp dualism between *puruṣa* (self) and *prakṛti* (matter, nature). The articles by Larson and Schweizer try to capture more precisely what this dualism amounts to, given that it is not equivalent to a Western mind-body dualism. Larson sees it as a kind of dualistic reductionism, Schweizer as a mind-consciousness dualism.

Advaita monism represents the most radical version of the *ātman* theory, according to which there is but one self which is identical with the Absolute *Brahman* (Deutsch 1969, Potter 1981). Deutsch's paper presents a "phenomenological" reconstruction of the Advaitin analysis of the self into a hierarchy of four levels of consciousness. Bhattacharyya's difficult but original piece presents a modern neo-Vedāntin perspective, influenced by Kant as well as the Sanskritic tradition (see also Bhattacharyya 1976).

References

Bhaduri, Sadananda (1975), *Studies in Nyāya-Vaiśeṣika Metaphysics*. 2nd ed. Poona: Bhandarkar Oriental Research Institute.

Bhartiya, M.C. (1973), *Causation in Indian Philosophy*. Ghaziabad: Vimal Prakashan.

Bhattacharyya, K.C. (1976), *Search for the Absolute in Neo-Vedānta*. Honolulu: University Press of Hawaii.

Chakrabarti, Arindam (1997), *Denying Existence*. Dordrecht: Kluwer.

Chakrabarti, Kisor Kumar (1999), *Classical Indian Philosophy of Mind: The Nyāya Dualist Tradition*. Albany: State University of New York Press.

Chattopadhyaya, Debiprasad and Gangopadhyaya, Mrinal Kanti (eds.) (1994), *Cārvāka/Lokāyata*. New Delhi: Indian Council of Philosophical Research.

Collins, Steven (1982), *Selfless Persons: Imagery and Thought in Theravāda Buddhism*. Cambridge: Cambridge University Press.

Deutsch, Eliot (1969), *Advaita Vedānta: A Philosophical Reconstruction*. Honolulu: East-West Center Press.

Dravid, Raja Ram (1972), *The Problem of Universals in Indian Philosophy*. Delhi: Motilal Banarsidass.

Garfield, Jay L. (1995), *The Fundamental Wisdom of the Middle Way: Nāgārjuna's Mūlamadhyamakakārikā*. New York: Oxford University Press.

Halbfass, Wilhelm (1992), *On Being and What There Is: Classical Vaiśeṣika and the History of Indian Ontology*. Albany: State University of New York Press.

Krishna, Daya *et al.* (eds.) (1991) *Saṃvāda: A Dialogue Between Two Philosophical Traditions*. Delhi: Motilal Banarsidass.

Larson, Gerald James (1979), *Classical Sāṃkhya*. 2nd rev.ed. Delhi: Motilal Banarsidass.

Larson, Gerald James and Bhattacharya, Ram Shankar (eds.) (1987), *Sāṃkhya: A Dualist Tradition in Indian Philosophy*. Princeton: Princeton University Press.

Matilal, Bimal Krishna (1968), *The Navya-Nyāya Doctrine of Negation*. Cambridge, Mass.: Harvard University Press.

Mohanty, J.N. (1992), *Reason and Tradition in Indian Thought*. Oxford: Clarendon Press.

Mukhopadhyay, Pradyot Kumar (1984), *Indian Realism: A Rigorous Descriptive Metaphysics*. Calcutta: K.P. Bagchi.

Parfit, Derek (1984), *Reasons and Persons*. Oxford: Oxford University Press.

Phillips, Stephen H. (1995), *Classical Indian Metaphysics*. Chicago: Open Court.

Potter, Karl H. (1963), *Presuppositions of India's Philosophies*. Englewood Cliffs: Prentice-Hall.

Potter, Karl H. (ed.) (1977), *Indian Metaphysics and Epistemology: The Tradition of Nyāya-Vaiśeṣika up to Gaṅgeśa*. Princeton: Princeton University Press.

Potter, Karl H. (ed.) (1981), *Advaita Vedānta up to Śaṃkara and His Pupils*. Princeton: Princeton University Press.

Sankaranarayanan, S. (1993), "The Pandit World with Special Reference to the Problems of Logic" in Guttorm Fløistad (ed.), *Contemporary Philosophy: A New Survey. Vol.7: Asian Philosophy*. Dordrecht: Kluwer.

Shastri, Dharmendra Nath (1964), *Critique of Indian Realism*. Agra: Agra University.

Shaw, J.L. (1978), "Negation and the Buddhist Theory of Meaning" *Journal of Indian Philosophy* 6: 59–77.

Smart, Ninian (1964), *Doctrine and Argument in Indian Philosophy*. London: George Allen and Unwin.

B. K. MATILAL

ONTOLOGICAL PROBLEMS IN NYĀYA, BUDDHISM AND JAINISM
A COMPARATIVE ANALYSIS*

The term 'ontology' came to be used to indicate the most general part of metaphysics in seventeenth-century Europe, although for the origin of ontology as a general theory of real entities, or as a theory of being *as* being, one has to go back to Aristotle as well as to the pre-Socratic philosophers of ancient Greece. Aristotle did not use the term 'ontology' just as he did not use the term 'logic' either. But the history of logic as well as ontology in the Western tradition seems to start with him. Aristotle talks about a 'first philosophy', which, he says, is about being as being, and this is taken in later Western tradition to be the nearest analogue of 'ontology'. For our present discussion, I shall assume ontology to mean a general theory of 'what there is' and try to apply it to the Indian tradition. There are many other problems usually discussed in connection with ontology in the West, such as the doctrine of distinction of essence and existence, the theory of transcendental properties of all entities, but these questions will not directly concern us in this paper.

The Nyāya-Vaiśeṣika ontological problem was connected with the Vaiśeṣika doctrine of categories (*padārtha*), and the category of substance was the focal point of this doctrine. The system of Vaiśeṣika categories is generally regarded as a classification of real and fundamental entities. It is also possible to view it as an analysis of the 'concrete' objects of our experience into their various parts in order to form a theoretical basis for our philosophical discussion. The Nyāya-Vaiśeṣika philosophers, however, believe that if we can analyse and classify the concrete object of our experience in this manner into substance, quality and action, we would achieve a satisfactory explanation of 'what there is', i.e., an explanation of what is meant when we say, 'that object exists'.

The Buddhists, on the other hand, think that the so-called concrete object of our experience is at best a synthetic object and hence is analysable into a number of fundamental properties or elements called *dharmas*. The Buddhist conception of a *dharma* is that it is by nature a non-substance (cf. *anātman*).[1] The question, 'What is there?' can be answered, according to the Buddhists, if we can prepare a satisfactory list of such non-substances or *dharmas*, which

we can refer to while we are accounting for and analysing the objects of our experience. The *dharmas* are also in perpetual flux, 'in a beginningless state of commotion', and *nirvāṇa* is posited as the ultimate cessation of this 'commotion' for a person. Nirvāṇa is also said to be the ultimate reality, the ultimate nature of things, to be contrasted with the phenomenal existence of *dharmas*, but, as I have already indicated, this problem will not be our concern in this context.

Our prephilosophical common-sense tells that there are around us things which somehow undergo change. Our philosophical worries start along with our recognition of the phenomenon of change vis-a-vis our feeling for continuity and sameness underlying change. In India this problem was reflected in the old dispute over *Sat*-cosmogony versus *Asat*-cosmogony (found in the Ṛgveda as well as in the Upanishads). The philosophic resolution of this dispute is to be found in the two rival theories about causation and creation in ancient India: 1. *sat-kārya-vāda* 'the theory of pre-existence of the effect in the cause' and 2. *asat-kārya-vāda* 'the theory of new creation of the effect which was non-existent before'. For those who prefer a comparative approach, it is significant to note that the so-called paradox of change and permanence, of being and becoming, was as much a live issue for the early Indian philosophers as it was for the Greeks, i.e., the pre-Socratics. Those who were inclined toward permanence not only posited the notion of an enduring substance but also argued that change was only superficial transformation of the existent (the substantial) from one state to another. The Sāṃkhya and the early Vedānta belonged to this group insofar as they gave prominence to *Sat* 'the existent'. The Vaiśeṣikas belonged to the group of *Asat* cosmologist inasmuch as they admitted change to be real and the function of the cause to be the creation of new things, effects. But they also posited the doctrine of substance, in fact, plurality of substance, and their substantial elements were said to be persistent through changing states. The Buddhist *Asat* cosmologists were very radical, for they argued that change alone was real and the notion of continuity or persistence was illusory, and the notion of soul-substance was a myth. The Jaina theory, as we will see later on, was a compromise between the Buddhists and the Nyāya-Vaiśeṣika.

The ontological positions of Nyāya-Vaiśeṣika, Buddhists and Jainas were necessarily influenced by their respective stands on the problem of change and continuity. The Buddhists, for example, were pre-eminently anti-substantialists in the Indian tradition. This anti-substantialism culminates in

their 'flux' doctrine, according to which the components of every object, all *dharmas*, change completely from moment to moment. A comparativist might be reminded here of the anti-substantialism of Heraclitus of Ephesus, who held, contrary to Parmenides' denial of change, that change was incessantly occurring. But it is not certain that the Heraclitean acceptance of change as reality amounted to the 'flux' doctrine, as it was understood by both Plato[2] and Aristotle.[3] The 'flux' doctrine may be due to an interpretation of the Heraclitean position by the philosopher Cratylus. This would at least give credence to the anecdote related by Aristotle about the 'river' example of Heraclitus. Aristotle says that Cratylus "rebuked Heraclitus for saying that you could not step twice into the same river; he (Cratylus) thought you could not even do so once."[4] Thus, probably Cratylus was much closer to the Buddhists in this regard.

The Nyāya-Vaiśeṣikas, on the other hand, were substantialists while they accepted also change much in the same manner as Aristotle. But we need not proceed in this comparative vein any further. It is important to understand now the Nyāya-Vaiśeṣika doctrine of existence as well as their notion of substance. Vaiśeṣika-sutra 8.14 asserts that what exists can be analysed into three categories, substance quality and action.[5] Existence in this system is regarded as a generic property common to the members of the three classes, substance, quality and action. Each of these classes has a class-property or generic property, viz., substance-ness, quality-ness and action-ness; but these generic properties are to be distinguished from 'existence' as a generic property.[6] Candramati, in fact, regarded existence as a separate category (*padārtha*) altogether while class-properties like substance-ness were included under the category of 'generality' (cf. *sāmānya* or *sāmānya-viśeṣa*). But Praśastapāda interpreted 'existence' as the highest generic property and thus brought both existence, on the one hand, and other class-properties like substance-ness and quality-ness, on the other hand, under one category (*padārtha*) called generality.[7] But still a special place was accorded to existence as the all-inclusive generic property which should be distinct from the included (*vyāpya*) generic properties such as substance-ness and quality-ness. A particular substance is characterized by the *being* of substance or substance-ness much as it is also characterized by many qualities and probably by some actions. But it is also characterized by 'existence' (inasmuch as it exists) which is not to be identified with its substance-ness or with any of its quality.

The best way to explain the notion of existence in this system is to

contrast it with the notion of 'real' as well as with that of the non-existent. Existence along with other included generic properties are themselves REAL but not EXISTENT. For, otherwise, we will have to indulge in talking about the existence of existence and so on *ad infinitum*. Similarly, the important relation called *samavāya* that combines the generic property existence with the particular existents, such as a substance or a quality or an action, is also regarded as REAL but not EXISTENT.[8] Thus, the generic properties and their inseparable relation with the particulars are posited as real, as means of explicating the notion of existence. Hence they themselves should not be construed as 'existents' to avoid the problem of self-dependence and regress but nevertheless, these notions, existence, generalities and *samavāya*, are claimed in this system to be real in the sense of their being independent of our thought or mind and thus being distinct from a non-entity. A non-entity is NON-EXISTENT and hence unreal, for example, the sky flower, the son of a barren woman, the rabbit's horn and the unicorn.

Briefly stated, the 'Existents' in this system (early Vaiśeṣika) are equivalent to the particulars, such as a chair, a particular color, a particular action. The class of existents is a sub-class of the class of reals. Universals (including relations) are not thus EXISTENTS but REALS. Praśastapāda used two significant notions in order to separate the class of existents from that of universals: *sattā-sambandha* and *svātmasattva*. The first notion characterizes each existent, for it means that EXISTENCE resides in the particular entity by *samavāya* relation. The second notion became a bit puzzling for the later commentators. Udayana explains it as 'lacking existence' (*sattā-viraha*). Śrīdhara gives almost the same interpretation but also points out that 'existence' could be ascribed to the universals only by mistake. Vyomaśiva says that while the first notion means that EXISTENCE is correctly applied to the class of particulars, the second notion means that EXISTENCE is only metaphorically applied to the class of universals.

The riddle of existence and non-existence is further complicated in the later Nyāya-Vaiśeṣika by the acceptance of negative properties as real. Our negative statements, according to the later Nyāya-Vaiśeṣika, are expressions of something, some negative facts. The affirmative-negative dichotomy among judgments is interpreted differently in this system. Thus, just as a positive judgment attributes a positive property to a thing so also a negative judgment attributes another property, a negative one, to the thing denoted by the subject-term. Just as a positive property predicated by a judgment can be

construed as a real property, so also a negative property, absence of some positive property, predicated by a judgment can be construed as a real property. Thus, 'the room is dark' can be interpreted as expressing the room that is characterized by the property of absence of light. Now, this property, absence of light, and the like, are regarded by the Nyāya-Vaiśeṣika as REAL inasmuch as they are to be distinguished from the unreal such as the round square, and unicornhood. But again, care should be taken to note that the absence of light is not however EXISTENT in this system in the sense a substance or a quality or an action is existent. It may also be noted that although the negation of an entity is construed in this system as expressing absence of that entity, a so-called negative property, no non-entity like the sky flower or the unicorn can be negated (in other words, absence of such non-entities will not be an acceptable negative property in this system).[10]

Leaving aside the riddle of existence and non-existence, let us concentrate on the Nyāya-Vaiśeṣika doctrine of substance and quality, which was at the focal point of their ontology. Several notions of substance have been emphasized in the Vaiśeṣika at some time or other: 1. substance as the locus of qualities and actions,[11] 2. substance as the substratum of change,[12] and 3. substance as capable of independent existence.[13] It is difficult to say whether the concept of substance as the logical subject was at all implied in early Nyāya-Vaiśeṣika doctrine, for it was never thoroughly worked out. Later Nyāya and Buddhist logicians (notably Diṅnāga) developed the concept of *dharmin* 'property-possessor' which was the nearest Indian analogue for 'logical subject'. But this concept was regarded as neutral to the ontological beliefs of the logicians. The concept of substance as the unchanging 'essence' was prevalent in the Sāṃkhya school as well as in the early Vedānta (cf. the spiritual substance), but this concept was not treated seriously in the Vaiśeṣika school. It is also to be noted that the Mādhyamika Buddhists were uncompromising critics of the doctrine of *sva-bhāva* 'own-nature' which was analogous to the notion of essence or inner immutable core of things.

The doctrine of substance as the substratum of change needs further elaboration in the present context, for this will throw much light on the Vaiśeṣika theory of causation and change. For any effect, the Nyāya-Vaiśeṣika will identify a particular substratum cause (*samavāyi-kāraṇa*) in which that particular effect is supposed to inhere. If the physical conjunction of two material bodies are taken to be the effect in question, its substratum cause will be the two bodies themselves. If the taste of a fruit is regarded as the

effect, its substratum cause will be the fruit-stuff itself. But when the effect is nothing but a concrete individual like a pot, its substratum cause will be the pot parts, or in final analysis, the atomic constituents of the pot material. Thus, the substratum cause of an effect need not be an ever unchanging substratum. We do not have to posit an unchanging substantial core as the locus of change. What is needed is only the temporal stability, persistence through a period of time, of the substance which acts as the locus of the effect.

The substances are, according to the Nyāya-Vaiśeṣika, either impermanent (having origin, stability and decay) or permanent (without origin or decay).[14] Material bodies of intermediate size (called *avayavin* 'whole' in this system) like a pot or a table are of the first type. They have temporal stability and can be the loci of qualitative change. These substances are breakable into parts and those parts into further parts. But the atomic constituents of these substances along with other non-material substances such as soul, sky, time and space are of the second type, i.e., permanent. An important part of this doctrine of substance is the ontology of the 'whole' (*avayavin*) as distinct from the assemblage of parts. A material body, e.g., a piece of chalk, is a whole which is a distinctly existent entity to be distinguished from the integration of its parts or combination of its atomic constituents. It is a new entity that is created as soon as the parts or atoms are put together. Moreover, draw a line on the board with this piece of chalk and you have created a new piece, for some parts of the old one are lost. The seeming identity of the new one with the old piece works for all our practical purposes, but ontologically the two are distinguishable.

The Buddhist anti-substantialism finds its extreme expression in the Sautrāntika doctrine of momentariness. According to this doctrine, a seemingly stable object like a chair is dissolved into a cluster of continuously fluctuating chair-moments or chair-stages. The real entity is a point-instant, an exclusive particular, an essentially unqualifiable, ineffable 'here-now' subject. Everything else in this system is only a conceptual construction – an interplay of the commonly shared imagination. In what sense does a moment exist? A moment exists insofar as it functions in some way or other. Thus, Dharmakīrti has argued that *to be* means to be capable of functioning in some way or other.[15] If a thing does not have causal efficacy, it does not exist. Starting from this initial position, Dharmakīrti and his followers have formulated a proof of their 'flux' doctrine. It will be interesting to note the crucial steps taken by the Buddhists in proving the 'flux' doctrine:[16]

1. To be is to do something, i.e., to function or to have causal potency.
2. To have causal potency means to be actually doing what is supposed to be done.
3. If something has causal potency at a particular moment it must do its work at that moment. (This is a rephrasing of 2.)
4. If something does not do a work at a given moment, it must be causally *impotent* to do that work. (This is a contraposition of 3.)
5. The same thing cannot be both causally potent at one moment and impotent at another (next) moment, for potency and impotency are contradictory properties, mutually incompatible.
6. Therefore, the thing at the moment of its potency must be held ontologically different from the thing at the moment of its impotency. A difference in qualities implies difference in the thing itself!
7. Everything, in this manner, can be shown to be in perpetual flux. We cannot step twice into the same river!

The most crucial step is taken by the Buddhist here when he identifies causal potency with actuality or actual doing. In other words, the notion of potentiality is completely rejected. If a thing exists and it is capable, it must function without lying in wait for anything to come and help. If we posit two different functionings at two different moments, we have to construe them as belonging to two different things or objects. In each moment a new object (*bhāva*) emerges when a new functioning sets in and the old functioning perishes. Thus, what exists is the ever fluctuating here-and-now. Even the ontology of stages or moments is not quite satisfactory to the Buddhists. For moments or stages are also hypothetical abstractions in the face of continuum. Thus, we have to say that there is only process, only flux, without something being there to fluctuate. There is only transmigration without there being any transmigrating soul (cf. the 'non-soul' doctrine).

Udayana, setting forth a defence for the Nyāya-Vaiśeṣika doctrine of substance, has criticized the above argument of the Buddhist by pointing out that it is essentially dependent upon the total rejection of the notion of potentiality. Why, asks Udayana, is it to be assumed that the causally potent cannot (and should not) 'wait' for its accessories. Causality operates with two mutually compatible notions: *svarūpayogyatā* 'potentiality' and *phalopadhāyakatā* 'actuality'; the former relates to the general while the latter relates to the particular. If the Buddhist equates potentiality with actuality then,

Udayana argues, part of the Buddhist argument is reduced to tautology, for he would have to say that x is actually functioning because it actually functions. And tautology is not a good philosophic argument. In fact, potentiality is explained by Udayana not as an essential constituent of the thing, but as the *mere presence of the thing coupled with the absence of some accessory or other and the consequent absence (or non-arising) of the effect.*[17] Thus, Udayana argues, if x does not cause y when and only when z is absent then it follows that when z is present x produces y. This is only another way saying that z is an accessory to x in bringing about y. Besides, the properties of causing y and not causing y are not two mutually incompatible characters like cow-ness and horse-ness. A cow, of course, can never be a horse. But a thing, if it is not just a flux, can cause y at time t_1 and may not cause y at time t_2.[18] In fact, what Udayana says is reminiscent of Aristotle's rejection of potentiality:[19]

> There are some who say, the Megaric school does, that a thing 'can act only when it is acting, and when it is not acting it 'cannot' act, e.g. that he who is not building cannot build, but only he who is building, when he is building; and so in all other cases. It is not hard to see the absurdities that attend this view.

It is rather significant that the arguments and counterarguments of Dharmakīrti and Udayana were presupposed much earlier in the Megaric school as well as in Aristotle.

Dharmakīrti's argument to prove his 'flux' doctrine was not entirely an innovation in Buddhist tradition. He must have derived his idea from Nāgārjuna's dialectic. Nāgārjuna, for example, has argued that if something exists it should exist always, and if it does not exist at one time it cannot exist at any time.[20] This is how Nāgārjuna has criticized the concept of existence and 'own-nature'. Dharmakīrti first assumes that *to be* means to have causal potency. Then he argues: if something has causal potency it must be functioning all the time, and if something does not have the causal potency at one time it would never have it at any other time. But Nāgārjuna's philosophic conclusion is rather different from that of Dharmakīrti. With the above argument Nāgārjuna wishes to avoid the extremes of eternalism and annihilationism and follow the Middle Way. Dharmakīrti, on the other hand, intends to conclude that since functioning is instantaneous, existence is also instantaneous. And when we think of Udayana's counterargument, we are again reminded of Aristotle:[21]

Again, if that which is deprived of potency is incapable, that which is not happening will be incapable of happening; but he who says of that which is incapable of happening either that it is or that it will be will say what is untrue; . . . . But we cannot say this, so that evidently potency and actuality are different (but these views make potency and actuality the same, and so it is no small thing they are seeking to annihilate), ...

The Jaina ontological position is influenced by both the Buddhists on the one hand and the Nyāya-Vaiśeṣika on the other. The Jainas were also substantialist, but in a very qualified sense of the term. Their conception of existence (*sat*) is intimately related to their doctrine of substance. The *Tattvārthasūtra* 5.29 asserts: "What there is, has the nature of substance."[22] And the next sūtra (5.30 in the Digambara tradition) adds: "What there is (the existent), is endowed with the triple character, origin, decay and stability (persistence)."[23] The *Tattvārthabhāṣya* explains that whatever originates, perishes and continues to be is called the existent; anything different is called the non-existent.[24] The next sūtra asserts that the existent is constant for it never gives up its being (essence?).

In sūtra 5.37, the substance is again characterized as the possessor of qualities (*guṇa*) and modes (*paryāya*). Here the broad category 'attribute' is apparently broken into two sub-categories, qualities and modes. But the distinction between qualities and modes is not found in the sūtra. Umāsvāti points out that qualities are permanent attributes of the substance while the modes are only temporary attributes which are subject to origin and decay.[25]

In the above analysis of the *Tattvārthasūtra*, two compatible notions of substance are emphasized: 1. substance as the core of change or flux and 2. substance as the substratum of attributes. Kundakunda combines these two notions as he defines substance in his *Pravacanasāra*:[26]

They call it a substance, which is characterized by origin, persistence and decay, without changing its 'own-nature', and which is endowed with qualities and accompanied by modifications. For the 'own-nature' of the substance is its existence (*sad-bhāva*) which is always accompanied by qualities and variegated modes, and at the same time, by origin, decay and continuity.

The Vaiśeṣika school, as we have seen already, emphasized both these aspects of substance, but did not equate the 'own-nature' of the substance with EXISTENCE. Aristotle, who in fact suggested several notions of substance either implicitly or explicitly, remarked in *Categories*:[27]

The most distinctive mark of substance appears to be that, while remaining numerically one and the same, it is capable of admitting contrary qualities.

In *Metaphysics*, Aristotle also implied that the substance is what is independently existent, for existence, in the proper sense of the term, applies to substances only, and qualities and relation have a secondary existence, a parasitic mode of being:[28]

Therefore, that which *is* primarily, i.e., not in a qualified sense but without qualification, must be substance.

The Jainas too, identify the notion of 'it is' or 'it exists' with that of substance, and they then explain that 'it is' means that it is endowed with the triple character of origin, decay and stability.

In fact, the Jainas explicated the notion of substance in such a way as to avoid falling between the two stools of being and becoming. It was a grand compromise of flux and permanence. The Jainas inherited from Mahāvīra and his later followers the well-known doctrine of 'many-natured' reality (cf. *anekānta-vāda*), and thus a 'compromise' position was only an important trait of their creed. The substance, in their analysis, is being, it is also becoming. Kundakunda observes that a substance has both natures: from the standpoint of one 'one-nature' it is being (*sat*, unchanging), and from another standpoint it has triple character, origin, decay and continuity, i.e., fluctuations.[29] Siddhasena Divākara repeated the point more forcefully:[30]

There is no substance that is devoid of modification, nor is there any modification without an abiding something, a substance. For origin, decay and continuance are the three constituents of a substance.

It should be noted that the notion of continuity involved in the triple character of the substance is not identical with the notion of permanence of the substance. The former notion means persistence or continuance (cf. *pravāha-nityatā*). The later notion means immutability. It is the notion in the background of which the triple character of origination, destruction and continuity becomes understandable. The notion of continuity, on the other hand, is essentially dependent upon origin and decay. Thus, Kundakunda observes:[31]

There is no origin without destruction, nor is there any destruction without origin, and neither destruction nor origination are possible without what continues to be.

The Jainas were well aware of the Mādhyamika critique of the 'own-nature' concept as well as the problem involved in the doctrine of the permanent substance. It is true that the immutability of own-nature invites a host of problems. But the notion of flux, the Jainas points out, is not sacrosanct. Thus, just as the Buddhists argue that there is only fluctuation, there being no

permanent being, the Jainas take the bull by the horns and answer that if there is no permanence there cannot be any change, any fluctuation, for it is only the permanent that can change. It is only the persisting soul that can transmigrate!

When the *Tattvārthasūtras* defines substance as the substratum of qualities and modes, it was probably influenced by the Vaiśeṣika school. Thus, Siddhasena points out that the rigid Vaiśeṣika concepts of substance and quality were not compatible with the Jaina ontological principle of *anekāntatā* 'many-naturedness' or 'non-onesidedness'. In fact it would be as good as a heresy in Jainism if one intends to maintain a rigid distinction between substance and quality. The notion of triple character, origin, decay and continuity, embodying the principle of (conditioned) reality, was derived from the Buddhist source. The Buddha, for example, predicated this triple character of all the conditioned (*saṃskṛta*) entities. Thus, in the *Aṅguttara* I, the Buddha said:[32]

> Of the conditioned entities, monks, the origin is conceived, even so their decay and their stability (persistence).

Nāgārjuna, however, directed his dialectical attack against the notion of the conditioned (*saṃskṛta*), and concluded:

> Since the notion of origin, persistence and decay cannot be established, the conditioned does not exist. And if the conditioned is not established, how will the unconditioned be established?

But why then did the Buddha speak about the triple character of the conditioned entities? Nāgārjuna replied:[33]

> Just as magic, dream and the cloud-castle are unreal (but, nevertheless, are spoken about) so also origin, stability and decay have been described.

The Jainas postulate the triple character in the case of each event, each happening or change of state. Each fluctuation embodies origin, continuity and decay. Samantabhadra illustrates the point as follows: if a golden pot is destroyed and a golden crown is made out of it, destruction, origination and continuity – all three – happen simultaneously and give rise to sorrow, joy and indifferent attitude in the minds of three different kinds of people, those in favour of the pot, those in favour of the crown, and those in favour of the gold stuff.[34]

Siddhasena has shown great philosophic insight in expounding the Jaina ontological problem. According to him, reality can be viewed from two

important standpoints, being and becoming, permanence and change. That is why Lord Mahāvīra acknowledged only two *nayas* or standpoints: 'substance exists' and 'modification exists'. If x is an element of reality, then, according to Siddhasena, x can be viewed as a SUBSTANCE from the standpoint of being, and as a PROPERTY from the standpoint of becoming. The standpoint of 'becoming' (modification) reveals that everything originates, stays and perishes; the standpoint of 'being' ('it is') reveals everything as existent, eternally without birth or decay. And, Siddhasena asserts, there cannot be being without becoming, or becoming without being; therefore, a substance (= reality) is defined as the combination of being (the existent) with becoming (origin, stability and decay).[35]

The 'being' aspect is, according to Siddhasena, the result of generalization while the 'becoming' aspect is that of particularization. In our ordinary description of things, we necessarily combine the general with the particular. From the point of view of the highest generalization, a thing is described as 'it is' which reveals the permanent being, the substance. But when, in ordinary descriptions, a thing is called a piece of wood, or a chair, or a red chair, we have an intermixture of 'being' and 'becoming' aspects. Insofar as the thing is identified as a non-fluctuating substance, it is the 'being' standpoint. And insofar as the attributes of the thing, such as being a piece of wood, being a chair, or redness, are revealed by the description, it is the 'becoming' standpoint. Qualities are nothing but modes or states of the substance. In any characterization or description of the thing there is thus an overlap of 'being' and 'becoming' standpoints, until we reach the ultimate particularity, pure BECOMING, i.e., the point-instants (*kṣaṇas*) of the Buddhists.[36]

Thus, the Jaina conception of reality, in bringing together the opposing viewpoints of the Buddhists and the Nyāya-Vaiśeṣika, comes very close to that of Whitehead, according to whom the chief aim of philosophy is the "elucidation of our integral experience", of both the flux and permanence of things. Whitehead has said that philosophers who have started with 'being' have given us the metaphysics of 'substance' and those who have started with 'becoming' have developed the metaphysics of flux. But Whitehead point out the inseparability of the two:[37]

> But, in truth, the two lines cannot be torn apart in this way; and we find that a wavering balance between the two is a characteristic of the greater number of philosophers. Plato found his permanences in a static, spiritual heaven, and his flux in the entanglement of his forms amid the fluent imperfections of the physical world. . . . Aristotle corrected

his Platonism into a somewhat different balance. He was the apostle of 'substance and attribute', and of the classificatory logic which this notion suggests.

In the Indian context, one may observe that the Buddha's search for *nirvāṇa*, the unconditioned state, freedom from suffering or *duḥkha*, spelled out a philosophy for the later Buddhists, according to which the flux of things, impermanences, "the fluent imperfections of the physical world" are identical with suffering (*duḥkha*, cf. whatever is impermanent, is suffering). And NIRVĀṆA, the unconditioned state, is actualized with the cessation of this *duḥkha*. The Vaiśeṣikas, on the other hand, were, much like Aristotle, the apostles of substance-and-attribute duality.

NOTES

* This paper was originally presented to panel discussion at an International Philosophy Conference held in New York, April 1976.

[1] Stcherbatsky, pp. 24–25.

[2] *Cratylus* 402a: "Socrates: Heraclitus is supposed to say that all things are in motion and nothing at rest; he compares them to the stream of a river, and says that you cannot go into the same river twice."

[3] *Metaphysics* 1010a: "It was this belief that blossomed into the most extreme of the views above mentioned, that of the professed Heracliteans, such as was held by Cratylus, . . ."

[4] Ibid., 1010a, l. 13–14.

[5] *Vaiśeṣika-sūtra* 8.14: *"artha iti dravya-guṇa-karmasu"*.

[6] Consult *Vaiśeṣika-sūtras* 1.2.7–1.2.18.

[7] Thus it is that Praśastapāda explains *sāmānya* as being of two types: *para* and *apara*. *Parasāmānya* is existence. Substance-ness etc. are *aparasāmānya*. See Praśastapāda, p. 15. For contrast, see Candramati, pp. 99–101.

[8] I am using the term REAL for the Vaiśeṣika term *padārtha*, EXISTENCE for the Vaiśeṣika term *sattā*.

[9] Praśastapāda, pp. 20–21; see also Udayana, p. 21; Śrīdhara, pp. 49–50; Vyomaśiva, pp. 118.

[10] See B. K. Matilal, 'Reference and Existence in Nyāya and Buddhism', *Journal of Indian Philosophy* 1, No. 1 (1970).

[11] Vaiśeṣika-sūtra, 1.1.7: *"kriyā-vad guṇa-vat samavāyi-kāraṇam iti dravyalakṣaṇam."*

[12] Vaiśeṣika-sūtra, 1.1.17: *"dravya-guṇa-karmaṇāṃ dravyaṃ kāraṇaṃ sāmānyam."*

[13] This is implied by the Navya-nyāya doctrine that a substance can exist by itself at the moment it is produced, without being joined by qualities and action (cf. *utpatti-kṣaṇāvacchinno ghaṭo nirguṇo niṣkriyaś ca*).

[14] See Praśastapāda, p. 22. Udayana comments: *"anitya-dravyatvaṃ cānyatra nir-avayava-dravyebhya iti."*

[15] *Pramāṇavārttika*, Ch. 2, v. 3: *"arthakriyāsamarthaṃ yat tad atra paramārthasat*, p. 100.

[16] For the most elaborate presentation of the 'flux' doctrine, see Jñānaśrīmitra, pp. 1–159.

[17] See the first chapter of *Ātmatattvaviveka*, specially pp. 16–25: "*yad yad-abhāva eva yan na karoti tat tat-sadbhāve tat karoty eveti.*"
[18] Ibid., p. 24.
[19] *Metaphysics*, 1046b, 28–32.
[20] *Mādhyamika-śāstra*, Ch. 7, verses 30–31.
[21] *Metaphysics*, 1047a, 10–20.
[22] "*Sat dravya-lakṣaṇam*", Tattvārthasūtra 5.29.
[23] "*Utpādavyayadhrauvyayuktaṃ sat*", ibid. 5.30.
[24] See Umāsvāti, under sūtra 5.29.
[25] Ibid., under sūtra 5.40.
[26] Kundakunda, *Pravacanasāra*, Ch. 2, verses 3 and 4.
[27] *Categories* 4a, 10–14.
[28] *Metaphysics* 1028a, 29–30.
[29] Kundakunda, *Pravacanasāra*, Ch. 2, verse 7.
[30] Siddhasena, *Sanmati*, Ch. 1, verse 12.
[31] Kundakunda, *Pravacanasāra*, Ch. 2, verse 8.
[32] Quoted by Candrakīrti, see Nāgārjuna, *Mādhyamika-śāstra*, p. 73.
[33] *Mādhyamika-śāstra*, Ch. 7, verse 34.
[34] Samantabhadra, *Āptamīmāṃsā*, Ch. 3, verse 57.
[35] Siddhasena, *Sanmati*, Ch. 1, verses 11 and 12.
[36] Ibid., Ch. 1, verse 9.
[37] A. N. Whitehead, *Process and Reality*, pp. 240–242.

BIBLIOGRAPHY

Aristotle. *The Basic Works of Aristotle* (ed. R. McKeon), Random House, New York, 1941.

Candramati. *The Vaiśeṣika Philosophy* (H. Ui). London, Royal Asiatic Society, 1917.

Dharmakīrti. *Pramāṇavārttika* (ed. Swami Dwarikadas Shastri), Bauddha Bharati, Varanasi (Benares), 1968.

Jñānaśrīmitra. *Jñānaśrīmitranibandhāvali* (ed. A. Thakur), K. P. Jayaswal Research Institute, Patna, 1959.

Kumārila. *Mīmāṃsāślokavārttika* (ed. R. S. Sastri Tailanga Manavali), Chowkhamba, Benares, 1898.

Kundakunda. *Pravacanasāra* (ed. A. N. Upadhye), Rajachandra Jaina Granthamālā, Agas, 1964.

Matilal, B. K. 'Reference and Existence in Nyāya and Buddhism', *Journal of Indian Philosophy* 1, No. 1 (October 1970).

Nāgārjuna. *Mādhyamika-śāstra* (ed. P. L. Vaidya), Mithila Institute, Darbhanga, 1960.

Plato. *The Collected Dialogues* (eds. E. Hamilton and H. Cairns), Bollingen Series, Pantheon Books, New York, 1963.

Praśastapāda. *Padārthadharmasaṃgraha*, with *Kiraṇāvalī* (ed. J. S. Jetly), Oriental Institute, Baroda, 1971.

Samantabhadra. *Āptamīmāṃsā* (ed. Gajadharlal Jain), Sanatana Jaina Grathamala, Benares (Kāśī), 1914.

Siddhasena Divākara. *Sanmati* (ed. Sukhlal Sanghavi and Bechardas Doshi), 5 pts., Gujarat Puratattva Mandir Granthavali, Ahmedabad, 1924–31.

Śrīdhara. *Nyāyakandalī* (ed. D. Jha), Varanaseya Sanskrit Vishvavidyalaya, Varanasi (Benares), 1963.
Stcherbatshy, Th. *The Central Conception of Buddhism* (Reprint), Matilal Banarssidas, Delhi, 1974.
Tattvārthasūtra (Umasvāmin) (ed. J. L. Jain), Bibliotheca Jainica, Arah (India), 1920. Reprint AMS Press, New York, 1974.
Udayana. *Kiraṇāvalī*. See Praśastapāda.
Udayana. *Ātmatattvaviveka* (ed. Dhundhirāja Śāstri), Chowkhamba, Benares, 1940.
Umāsvāti. *Tattvārthasūtra-bhāṣya* (ed. K. Siddhantasastri), Jaina Mandala, Bombay, 1932.
Vaiśeṣika-sūtras of Kaṇada. (ed. Muni Sri Jambuvijayaji), Oriental Institute, Baroda, 1961.
Whitehead, A. N. *Process and Reality – An Essay in Cosmology*. The Free Press, New York, 1969 (first published by Macmillan Co., 1929).

KARL H. POTTER

VEDĀNTAPARIBHĀṢĀ AS SYSTEMATIC RECONSTRUCTION

The *Vedāntaparibhāṣā* (VP)[1] of Dharmarājadhvarīndra (fl. c. 1615) is frequently indicated as an introductory handbook for the study of Advaita Vedānta, analogous to such works as the *Tarkasaṃgraha* for Nyāya, or the *Mīmāṃsāparibhāṣā* for Pūrvamīmāṃsā. To identify it in this way is, however, perhaps to miss the VP's most significant and remarkable feature, which is that it represents the most obvious case of philosophy as system construction found in Advaita literature. The significance and remarkability of this is that it gives the lie to a mistaken impression often disseminated about Advaita, which is that it is by the very nature of its thought mystical, unsystematic or even antisystematic. While those acquainted with Advaita literature are likely not to be victimized by this mistaken impression, it is a common enough one to be worth exposing, especially since it is sometimes implied that it is precisely this mystical, un- or anti-systematic nature which constitutes the glory of Advaita. The *Vedāntaparibhāṣā*, I shall attempt to show, is systematic in a rigorous sense of the word, and it is instructive to view Advaita in its terms, which stand at the culmination of a history of many centuries of careful work of developing Advaita as system.

When I say that the *Vedāntaparibhāṣā* is "systematic" in a rigorous sense of the word, what do I have in mind? In a general sense, a system is an ordered exposition. But there are various ways of achieving order. Dharmarāja's work, I wish to argue in this paper, is instructively viewed as an especially rigorous kind of ordered exposition, an exposition organized in terms of a series of interlocking definitions derived from a small number of primitive terms.

I

Dharmarāja's work turns around some 84 definitions. Definitions are, of course, not new to Advaita in the *Vedānta-paribhāṣā*: they are a regular part of Advaita, and of Indian philosophy generally, since its earliest beginnings. What makes Dharmarāja's definitions unusual is that they are interconnected with each other so that all of them can be generated from

[1] Dharmarāja Adhvarin, *Vedāntaparibhāṣā*, edited and translated by S.S. Suryanarayana Sastri. (The Adyar Library, Adyar, 1942). Hereafter abbreviated as "VP." References are to this edition.

a small base of primitive notions. This is what I shall attempt to set out—though necessarily only partially—in what follows.

But suppose that I can show that these definitions are connected in this way, why do I suppose that is an interesting result? One may well doubt whether any purpose other than an aesthetic desire for elegance is served by relating a lot of definitions in this way. In order to answer this doubt I need to indicate how system-construction can be seen as a most powerful, as well as elegant, way of carrying forward a philosophical program.

One of the most important books of 20th century philosophy, a work too little known, is Nelson Goodman's *The Structure of Appearance*.[2] In it, a certain philosophical program is carried forward through system-construction; furthermore, one finds in the pages of Goodman's book, as well as in other published pieces of his, valuable observations and arguments concerning the nature and value of system-construction in the manner of SA (as I shall call Goodman's book and the system it develops). While I cannot do justice in a single lecture to the depth and breadth of Goodman's analyses, I shall have SA very much in mind as I proceed.

Let us start from some truisms about what philosophy is, or may be, about. A philosopher, perhaps among many other things, sometimes attempts to clarify and explain a certain subject-matter in keeping with some purpose or set of purposes (such as clarification of meaning; discovery of truth) to which he is committed. Thus he might be thought of as a kind of mapmaker: he explores a territory and "maps" it in such a way as to enable those who use his map to find their way around in the territory assuming they are guided by the kind of purpose(s) he had in mind. So, a highway map of Italy helps a motorist to proceed from Naples to Rome, to identify towns near the main route if they wish to deviate for lunch, and so on. The relativity to purpose of a map is a point worth noticing. A mapmaker is not trying to recreate the territory: as Goodman says, "let no one accuse the cartographer of merciless reductionism if his map fails to turn green in the spring."[3] Recreating the domain is not what the mapmaker is up to.

In the case of highway mapmaking the domain is clear enough: it is provided by the features of the territory being mapped, those features which are relevant to the purposes in point, viz., getting from here to there most efficiently and by established roads, etc. But mapmaking is evidently an analogy: a philosopher is not a cartographer in the literal sense of the term. And so one may well ask: what is the domain the philosopher treats in a

[2] Nelson Goodman, *The Structure of Appearance*, (Harvard University Press, Cambridge, Mass., 1951; Second Edition 1966, Bobbs-Merrill; Third Edition 1977, D. Reidel). Hereafter abbreviated as "SA." Page references below are to the third edition.

[3] Nelson Goodman, *Problems and Projects* (Bobbs-Merrill Company: Indianapolis, 1972), p. 16.

fashion analogous to the way the cartographer treats a territory of space?

There is no single answer to that. Philosophers have had various purposes in mind, and in turn these picked out various aspects of things as worthy of their attention. But we may be able to devise a general way of speaking about the domains of philosophical explanation if we consider another way of thinking about what the mapmaker does. We could say that the highway mapmaker, considering a domain and having in mind serving a certain purpose, provides a method of giving a clear answer to all questions which a motorist might raise about his relative location to roads and towns and whatever else he needs to know to satisfy his purposes as motorist trying to get from one point in the domain to another. Thus, one might say, the map provides the motorist with a way of answering all such questions (perhaps within certain restrictions which are understood implicitly or explicitly). Or, to put it another way, if we consider all the things a motorist might say about his relative location *vis-à-vis* roads and towns (e.g., "I am going north from Naples," "Rome is 20 kilometers ahead," etc.), the map tells us which of those statements are true and which are not true, i.e., which ones to trust and act upon and which ones not to. Likewise, we can conceive of the philosopher providing a similar service for those oriented toward the purposes the philosopher has in mind serving. Whatever that purpose is, the philosopher will produce a translation manual in which, for every statement germane to that purpose (and within certain restrictions which are understood implicitly or explicitly) there either is a translation of that statement in the manual, or a translation of its contradictory, or neither, but not both. If there is a translation in the manual, that shows the statement to be true; if of its contradiction, that shows the statement to be false; and if neither, that shows the statement to be confused or irrelevant or that it conflates more than one statement, so that it is neither true nor false *simpliciter*, though it may or may not be related to statements in the manual in other ways. Once again, the set of statements in the manual provides a way of telling which of the original statements to trust, and which not.

We have, then, the conception of a system as a translation manual or scheme the input into which is a set of ordinary language statements and the output from which consists of translations of some of these together with clarification of the others. The limits on the input sentences reflects the philosopher's choice of the purposes he wishes to serve and the domain over which he wishes to range in his translations. Philosophy is frequently philosophy of something—philosophy of science, philosophy of mathematics, philosophy of art, political philosophy, and so on—and these descriptions identify in a general way what sorts of statements those kinds of philosophers are concerned to translate. Of course, "philosophy of science," etc., are merely very general rubrics, and we can hardly expect

them to classify neatly all the purposes and domains philosophers commit themselves to.

We have, then, the conception of a philosopher constructing a translation manual for a group of ordinary language statements picked out in reference to a purpose and a domain. But why should he do such a thing? What's wrong with the statements in the first place? Why do they need translating? The answer, of course, is that ordinary thinking about a domain and a purpose may be very fuzzy. Our language and concepts are frequently infected by vagueness and ambiguity, with the result that confusions arise, and we sometimes act on the basis of statements which are either flatly false or else so confused that they provide little or no basis for successful action. So, the translation manual should be written in a language which is, by comparison with ordinary language, precise, unambiguous, and whose statements are consistent with one another. Thus, just as a map is a bad one if its design is vague, its symbols ambiguous, and the directions as to how to proceed derived from it inconsistent with each other, a translation manual is a bad one if its syntax and/or terms are vague and ambiguous and if its constituent statements are mutually inconsistent, and for analogous reasons.

To put it in more formal terms, the criteria for a successful translation manual include internal (1) consistency with each other of the sentences in the manual and sufficient (2) clarity to enable a person using the manual to manage the domain and achieve the purpose delimiting the scope of the manual. To these we may add a few other criteria. First, the manual must be able, at least in principle, to translate *every* ordinary language statement falling within the limits of the domain and purpose. This criteria I shall call (3) adequacy. The manual must give us a workable basis for making decisions as to how to achieve the purpose, so that by acting on those decisions we succeed in achieving the purpose. This criterion I shall call (4) accuracy. And the manual must be intelligible yet powerful, which is to say that the audience for whom the manual is intended should be able to understand the terms in which the translations are couched as clearly derived from a small basic vocabulary with which he is, or can be made to be, familiar and which is itself sufficiently clear. This criterion I shall call (5) economy. To revert to the analogy of mapping in order to shed some further light on the last three criteria, it should be apparent that if a map only shows *some* features, say, only some roads relevant to the purpose, but not others it is less satisfactory than a map which shows *all* such features, i.e., it is relatively inadequate. A map which is misdrawn, so as to direct a person trying to get to Rome towards Florence instead, is inaccurate. And a document with no key to its symbols, or which uses a different symbol each time for recurrent features, is hardly worth calling a map.

To the extent that inconsistency, unclarity, inadequacy, inaccuracy and lack of economy infect a translation manual, to that extent it is less satisfactory than we might hope.

What might be the organization of a translation manual that would be most likely to satisfy, or best approximate to satisfying, these criteria? It should be written in a language that its intended users can understand, a language which is recognizably connected with the ordinary language in which the sentences to be translated are couched, so that it will be evident that those sentences are indeed being translated. Yet the point of the translation is not, of course, to reproduce the deficiencies (unclarity, inconsistency, etc.) inherent in the sentences to be translated. The translations in the manual must be in a "clarified" version of the language, one in which both the syntax and the semantics of the translation sentences are carefully controlled so as best to achieve satisfaction of the criteria mentioned. Goodman argues cogently that what this conception of the philosopher's task suggests is the construction of a *system* of *definitions*.

The translation manual should be organized around definitions. These definitions, notice, are definitions in the manual, i.e., they are not definitions of the concepts of the manual in terms of the concepts of the language to be translated. Nevertheless, since the language of the manual must be intelligible to users of the language being translated, this means that the definitions must ultimately be couched in terms understandable and explicable in the ordinary, presystematic language and so subject to whatever unclarity is inherent in those presystematic understandings and explications. For this reason alone, the fewer "primitive," that is, undefined terms that the definitions rest on, the fewer opportunities for confusion. By minimizing the occasions for confusion through depending on as few primitive terms as possible, we also minimize possible inconsistencies that might crop up if we failed to follow a rigorous and systematic way of proceeding from definition to definition. This suggests that the definitions should constitute a system, i.e., an interconnected group of translations involving the primitive and defined terms organized through the interconnections among those terms. In short, the procedure is to select a few primitive terms—as few as we can get away with—and then to define other terms in terms of those primitives, still others in terms of the stock of defined terms plus primitives at hand at a given point in the process, and so on until we have provided a vocabulary in which every sentence in the domain and relevant to the purpose finds a translation into that vocabulary.

That is the rationale for system-construction of the sort that Goodman practices in SA. He takes his cues to some extent from an earlier exercise

in system-construction. Rudolf Carnap's *Die Logische Aufbau der Welt*;[4] Indeed, Goodman provides an analysis and critique of Carnap's construction in SA itself.[5] Since I am more familiar with Goodman's own system, and since it is written in English, I shall spend a bit of time characterizing the system constructed in SA as an illustration of the sort of translation manual I have been talking about.

Goodman points out that in going about constructing a system one must initially make three kinds of choices. First, one must decide what sort of syntax, that is to say, what logical connectives one will use in forming the sentences of the translation manual. Goodman's own choice here is to use standard quantificational symbolic logic of the sort developed in *Principia Mathematica* and commonly taught in logic classes today in American and British universities. It might be noted that, in apparent contradiction to something I have just said, the language of the manual, symbolic logic, does not look like the language of the sentences to be translated into it, which is ordinary English. Goodman notes this but points out that there is no problem since every interpreted symbolic formula which appears in the manual can be rephrased into English; the advantage of the formula is clarity and rigor, the value of the English rephrasing is intelligibility, and they are both present, though requiring distinct representations, in the sentences of the manual as Goodman conceives them. Standard quantificational logic contains as elements logical connectives, variables and quantifiers, predicates and names. Goodman's plan is to represent the syntax of the sentences in his translation manual, i.e., his systematic translations, by using formulae of standard quantificational logic together with their English equivalent expressions.

The second choice a system-constructor has to make, Goodman tells us, concerns the ranges which the variables will be taken to cover and the kinds of things the names appearing in the formulae will be allowed to denote. The reason why this is an important choice is that it may be thought—and is so thought by Goodman—to represent an initial choice of an ontology. According to a well-known thesis, implicit in *Principia Mathematica* but regularly associated with the authority of W.V.O. Quine, "what there is," i.e., what one is ontologically committed to, is a matter of the values over which one's variables are allowed to range. This is clear from the meaning attached to the existential quantifier $\exists$, which is to be read as "there is at least one x such that," committing one to the existence of at least one thing of a certain kind. Since logical transformation can,

[4] Rudolf Carnap, *Die Logische Aufbau der Welt* (Berlin Schlachtensee: Weltkreis-Verlag, 1928). An English translation by Rolf A. George was published in 1967 from the University of California Press, Berkeley, California.

[5] SA, Chapter 5.

though under restrictions, lead from universally quantified statements to existentially quantified ones, the variables over which universal quantifiers are allowed to range also reflect ontic commitments. In short, the ranges of the values of the variables over which one's quantifiers extend provide one kind of primitive notion for the system. If we understand "(x) (Man x $\supset$ Mortal x)" as saying that for anything, if it is a man then it is mortal, we are making use of the primitive notion of a *thing*. If the translation manual contains such expressions as

"$(\exists x)(\exists t)(O_{xt})$,"

read "there is at least one thing and at least one time such that that thing occurs at that time," we are invoking times as well as things as values of our variables. Other likely candidates for values of variables include places (i.e., positions in a spatial array), events, persons, thoughts, atoms, physical objects, etc. By recognizing one or another of these as values one admits them as presumptively unproblematic parts of one's ontology.

The third kind of choice that must be made is that of the primitives of the system. A primitive notion is one which is not defined itself within the system, but in terms of which other notions receive their definitions. In order that the requirements of intelligibility be respected, primitive notions must be explicable, but their explication is not strictly part of the system itself, but rather in ordinary language. For example, in SA Goodman's basic primitive is represented as "W," standing for "with" and called "togetherness," explicated as the relation which links each pair in an experience of a color at a time at a place in one's visual field. Colors, places and times are values of variables in Goodman's system, and a color's being with a time (for example) is something that happens in our experience but is not defined within his system: we are supposed to be able to recognize that kind of experience and, through any help by discussion or pointing or whatever, be able to identify the relation intended.

In system-construction of this sort one aims at the simplest basis of primitives. This is the criterion of economy that we mentioned above, also called "simplicity." It is frequently supposed that economy or simplicity is an ancillary desideratum, only worth seeking for elegance sake after the other criteria of success have been satisfied. Goodman disagrees, and I quote.

> Most generally speaking, the purpose of constructing a system is to interrelate its predicates. The same purpose is served by reducing to a minimum the basis required. Every definition at once both increases the coherence of the system and diminishes the number of predicates that need be taken as primitive. Thus the motive for seeking economy is not mere concern for superficial neatness. To economize and to systematize are the same.[6]

[6] SA, p. 48.

Goodman develops a way of comparing the relative economies of different choices of primitives by examining their relative strength as constituted by a function of their explanatory power and the number of places the predicates have. The concern for the number of places in a predicate arises because if one merely measured simplicity in terms of the number of primitive predicates a system has one could easily collapse all the primitives into a single one with an enormous number of places. Since any system can do this all systems will turn out to be equally economical. The resulting problem—of specifying just which combinations of number of primitives and number of places is preferable, which less so—turns out to be more difficult than one might have thought.[7]

Just which predicates one chooses as primitives is, of course, a function of the purposes one is concerned to serve and the nature of the domain one is systematizing, which is to say, the kinds of ordinary language statements one plans to translate. Goodman sees this choice as a possibly neverending matter of trial and error. One tries one basis of primitives, sees how far he can get by building definitions and thus finding ways of translating, until he runs afoul of one of the five sorts of criteria of success I mentioned earlier. At this point one goes back and starts again at some choice point, making a different choice of basis, trying a different kind of definition.

Having made an initial choice of primitives—and of values for the variables and a logic—one is ready to start trying to build a system of definitions. Goodman's own system has as its purpose translating ordinary statements about the physical world into a manual in which only phenomenal predicates are used, i.e., it is a systematic way of furthering the philosophical program called phenomenalism. As we have seen, his variables range over colors, times of presentation and places in the visual field. His initial primitive is W, togetherness. In terms of W he first defines what it is to be a "quale," i.e., to be one or another of the three basic kinds of things recognized in the system—colors, times and places. Then he defines the notion of being a "complex," describable in English as an entity such that every two discrete parts of it are together. "Complex" is thus a defined term in the manual or system. From there he moves on to define a concretum as a complex that is not itself together with any quale or complex. A concretum is a color-spot-moment, that is, it is the appearance of a color at a specific time in a specific place in one's visual field. Next, he defines "(quale x) is a quality of (complex y)," and then considers ways of defining predicates of the sizes and shapes of things. And

[7] Goodman discusses the problem in SA, Chapter III and several articles, some of them collected in *Problems and Projects*, op. cit., 279–355.

so on. The eventual aim is to be able to translate ordinary language and scientific statements such as "at 10 o'clock the 5-foot square box moved 3 feet north" into the system so constructed, in which the only primitive notions are phenomenal.

In this way the system-constructor painstakingly works away at providing answers to one or another fundamental philosophical question, or usually, a whole nest of questions. In Goodman's case the nest includes the so-called "mind-body problem" about the relation of the physical world to experience, and thus implicitly involves the entire epistemological controversy between realism and idealism. He also, at the same time, has chosen "atoms of the system," i.e., the values of the variables, which are repeatable entities (viz., qualia—colors, times, places), and another aspect of the program is to translate what we standardly speak of in terms of nonrepeatable particulars, such as events, objects and such, into the language of repeatable qualia.

You may say that it seems a very difficult task to no very clear purpose. Why should one want to translate sentences about physical things into one about mental states? One way of answering is to say that one might if he wondered whether there are physical facts or things over and beyond mental ideas and sensations. To revert to the mapping analogy, it would be as if we wondered whether we need to have symbols for, say, north, south, east and west, the compass-points, on our map or whether we could dispense with directions and get the same result by having symbols for up, down, left and right on the map. We would presumably find we can't since we cannot guarantee the reader will not try to read the map upside down, so to speak. That is a rather trivial example, since it would be an odd thing to deem the notion of directions problematic and needing clarification in this fashion.

Another, and more insightful way of answering the question, is to point out that though there are things we think we know for sure about our mental life, we are by no means so confident about things said about physical facts and things, or about the relations between mental and physical things. That is to say, there are lots of ordinary statements of this sort made by ordinary people and even by scientists whose truth-value is not at all clear. Consider "nothing can be red and green all over at the same time," "every event has a cause," "my mind is in my head," and all the myriad puzzling propositions that philosophers have traditionally puzzled over. These are statements that need translating into a clarified language, if only to get rid of them in place of some statements we can confidently assign truth-values to.

Of course, a lot of statements are such that if a translation manual rendered them false we should burn the translation manual rather than

change our minds about the statement. These are the statements that reflect the context of the purposes which generates the inquiry at hand. If the manual, for example, that is constructed in order to decide, among other things, whether "nothing can be red and green all over" should be such as to make true the translation of the ordinary sentence "nothing can be red at all," we should lose interest in that manual for that purpose. This is an important point to recognize. Inquiry is always for a purpose, and the parameters of a problem are a function of the problem. Solutions to the problem must occur within those parameters.

An important and, I think, refreshing feature of the attitude of the philosophical system-maker, certainly as Goodman construes it, is its lack of dogmatism. One constructing a system should be very wary of making any unwarranted claims for his particular choices in starting a construction. The choices of logic, atoms of the system, and primitive predicates are viable only if the system that results succeeds, which is to say, only if the system so based satisfies the criteria of success that we have alluded to better than any other rival system. But the number of possible rivals that are pertinent to any given domain and purpose is indefinitely many, and there is little reason *a priori* to suppose that one set of choices is more likely to succeed than another. There is no shortcut to success in system-construction. Thus prior to actual long, hard work one does not have sufficient reason to assume that physicalism rather than phenomenalism, realism rather than nominalism, or any other metaphysical position is necessarily true. These metaphysical positions become hypotheses when viewed as choice points in system construction, tentative lines of solution which may turn out to be impossible to sustain. Thus if the purpose is to resolve the mind-body problem, say, translation of the sentences of the domain into physicalistic language is one line of solution, translation of those sentences into phenomenalistic language is another. Carnap's *Aufbau* is an attempt to work out a system which reflects the first choice or hypothesis; Goodman's SA is an attempt to work out a system which reflects the second. Goodman gives arguments to show how Carnap's particular system breaks down, but he does not conclude from them that physicalism is bankrupt as the basis for a successful system. There are or may be other choices of basis which are physicalistic in spirit and which will succeed where Carnap fails.

This undogmatic attitude seems to me to be refreshing in that it keeps lines of inquiry open by contrast with dogmatic claims about metaphysical theses which close (or attempt to close) all lines of inquiry other than those consistent with the favored thesis. It seems especially pertinent when we consider the context in which Indian philosophy operates. Here all parties are committed to the same purpose, which we might characterize as con-

struction of a system or translation manual into which all statements pertinent to the gaining of liberation can be translated in a manner satisfying our five criteria. Antecedent to inquiry there is no reason to suppose that realism is better suited to solving this problem than idealism, that physicalism is preferable to phenomenalism as a base, or that nominalistic translations will work out better than realistic ones. Viewed in this perspective, the various schools of Indian philosophy represent the beginnings of the construction of systems each of which adopts different bases, i.e., makes choices of logic, atoms and primitives which are distinct from the others, but none of which have any *a priori* reason to suppose that their choices will succeed where others will fail.

II

In the *Vedāntaparibhāṣā* we find many definitions—at least 84 of them by my count. The terms defined include most of the technical terms of the Advaita school's philosophy. My project in this lecture is to suggest how these definitions can be understood as constituting a system of the kind illustrated by the system of SA. That is, I shall suggest a choice of atoms of the system (values of the variables) and of primitive predicates from which, it seems to me, one can derive a number of Dharmarāja's definitions. I shall not derive them in precisely the order Dharmarāja presents them, nor do I claim that Dharmarāja presents them, nor do I claim that Dharmarāja himself, if he were here, would necessarily approve of the particular choices and order of things which I here propose. My point is rather that Advaita can be viewed as a systematic philosophy, and that Dharmarāja's actual definitions, whether he intended them that way or not, can plausibly be incorporated into a system of the kind I have in mind.

Why do I think this a worthwhile exercise? For one reason, because Advaita is often viewed as mysticism, about as far removed from system as one can imagine. Indeed, some Advaitins like to say that themselves about their school, that is, they like to characterize their school in anti-intellectualist fashion. There is certainly a strong flavor of this in some of Śaṃkara's writings, and one might well, reading both primary and secondary literature on Advaita, come to the conclusion that Advaita is philosophy in an entirely different sense from schools such as Nyāya and other *darśanas*, that its purpose is by negative dialectic to remove the intellectual source of ignorance which veils Brahman, which removal constitutes Self-knowledge and liberation. In much of post-Śaṃkara Advaita, however, the school is not viewed that way; rather, it is understood by its proponents as committed to the same purpose as the other *darśanas*, viz.,

to show us how to think about things so as to reach liberation, and this is understood as requiring a positive account of the nature of things, since "Self-knowledge" is taken to require understanding, not merely removal of a veil. To put it slightly differently, the mystical interpretation essentially thinks of Advaita as a dialectical method of causing liberation to occur without inculcating any positive account of things to replace the mistaken notions of the other viewpoints. The nonmystical interpretation agrees that wrong views must be eliminated for one to be liberated, but avers that right views are revealed when the wrong ones are eliminated, and that these are views which form a proper part of a systematic philosophy as much as those of any other school's thought. The mystical interpretation finds the notion of system-construction repugnant, but the nonmystical interpretation has no more reason than any other school to resist systematization, and therefore has an interest in reasons for thinking its system to be comparable to, and hopefully superior to, the systems that can be developed by rival schools.

A second thing which exploring Advaita as system may do is to convince skeptics who (as I once did) feel that the Advaitin's fundamental insight, though a positive one, is not capable of consistently rigorous formulation, convince them that it does indeed support systematic development.

A third thing is to get a feel for what the organization of Advaita theory actually comes to, to bring out which notions are central and which are peripheral. The exercise of systematizing Advaita forces one to confront this sort of question.

These several things which systematizing Advaita can achieve do not actually require completion of the system for their achievement, which is a good thing since space does not allow complete reconstruction of all Dharmarāja's definitions here. It also does not require my following the text blindly: I can use a bit of imagination, or rather, I can import some insights born from the study of other systems, such as the one Goodman builds in SA.

We shall suppose that we are to systematize the insights of Advaita along the lines outlined in the previous lecture. The first thing to do, therefore, is to propose which choices one plans to make about logic, atoms and primitives. I shall in this exposition pass over the question of the choice of logic: the logic I shall use will be standard quantification theory as set forth, e.g., in Quine's *Methods of Logic*,[8] but I hope to avoid any difficult problems of a purely logical nature. This leaves us with the choice of atoms and primitives, a choice in this case dictated to some extent, though not

[8] Willard van Orman Quine, *Methods of Logic* (Henry Holt: New York, 1950, 1959).

completely, by the terminology of Advaita and in particular of Dharmarāja himself.

I propose that we admit three kinds of values of variables, i.e., three sorts of atoms of the system. One of these is moments of time. I choose this way of handling time because to define time in terms of other notions is cumbersome, and we need to make references to times almost from the beginning: if I were to try to derive temporality from other considerations I would have no room to do anything else here. In an improved system born of a different choice it is possible times could be eliminated as values and temporality handled as a primitive or defined predicate or family of predicates.

The other two kinds of things I need to have my variables range over arc "beliefs" (*niścaya jñāna*) and their contents (*viṣaya*). The strategy is to provide a means of translating all the things that Advaitins want to say about Brahman and *māyā*, levels of reality, perception and the other means of knowing, and so on into a language whose sentences are all about times and beliefs and their contents. If Quine is right, this means that provided this system is the most successful manner of representing the Advaita translation manual, Advaitins are committed to the existence of times, beliefs and contents. This doesn't, however, mean that they are not committed to other things as well. Nor does it mean that these three kinds of things are more real or more existent than others, say, Brahman. It does mean that, supposing this system is successful, the Advaitin commits himself to these things for the purpose of talking about and thinking about things generally.

The values of our variables, then, range over beliefs, contents of beliefs, and moments of time. The third and last choice I need to make concerns the number and nature of the primitives of the system. I propose to try to make do for present purposes with three primitive predicates, each having three places.

The first predicate corresponds to the Sanskrit notion of *viṣayitā*. It is the relation which holds between a belief and its content and the time at which the belief occurs such that that belief is about that content at that time. I shall represent this as "Ajct." Thus, if at 10 o'clock a.m. on the 1st of June 1983 a certain belief J occurs which has the form "this is a pot," we can say that J is about a certain pot at that time (a content *c*). Notice that a "content," as here understood, is not necessarily an actual object; it is an intentional entity, what a belief is about, its content. If I am thinking of a pot, there is something—a content, that pot—that I am thinking of even though my thought may be confused or misguided and there is no pot where I think there is one.

The second primitive predicate is termed, in Sanskrit, *avacchinnatva*. It

comes to the same sort of relation that used to be indicated by the Latin word "*qua*" in an expression such as "Being *qua* Being." We might parse it as the relation which holds, when I consider at a certain time *t* a certain thing *x* under a certain description *y*, between *x* and *y* at *t*, that is, that at *t x* is limited by having the form of *y* for the purpose of my thinking about it. This relation of limitation has several varieties in Advaita: it may be the case that *y* is an essential property of *x*, so that when I consider *x* as a *y* I am considering it as having one of its necessary constituent features. Or it may in another case be that *x* doesn't always have *y*, so that by considering it under the description "*y*" I am identifying it by its having features which are not part of its nature. These differences will be defined in due course. In the symbolic renditions of my definitions which follow I represent "*x* is limited by (having the form of) *y* at *t* as "Q_{xyt}."

The third primitive predicate, which answers to the Sanskrit *āśritatva*, is a relation of being located at, or in, or on, something. For example, a belief may be located (may occur) in someone or something at a time. I shall represent "*x* is located in (at, on) *y* at *t*" as "L_{yxt}."

If these primitives seem rather mysterious when introduced out of the blue in this fashion, that is to be expected. Their utility is to be demonstrated in the process of building up the system of definitions. Naturally, being primitive, they themselves do not get any definitions in the system.

I have now assembled the materials to begin constructing the system. Within the limits of space I now produce a series of definitions utilizing the atoms and primitives. Some of these definitions are precisely those that Dharmarāja provides, though not in the same order as he presents them. Others are definitions of technical terms that Dharmarāja takes for granted in his exposition but which, given my method here, we need to derive from the materials in hand before we are allowed to use them. I have written them out in the handout in logical formulae, but will explain them in English as I go along.

I start with Dharmarāja's 25th definition, in which he explicates the notion of sublation (*bādha*). Simply put, a belief is said to be "sublated" at a time *t* just if its content is believed to exist up to time *t* but at that point the person whose awareness it is comes to disbelieve that that content exists, i.e., he ceases having that kind of belief. (Dharmarāja's definition = DD).

DD25: Bādha = kāryavināśa upādānena saha.[9]

I express this in *my* first definition. (Potter's definition = PD)

[9] VP, p. 46.

$$\text{PD1: } Bj_1t_{2+} = \text{df. } (\exists t_1)(\exists c)(\exists x)\ (Aj_1\,ct_1 \cdot Lxj_1t).\ (j)(\sim Ajct_{2+} \cdot \sim Lxjt_{2+})$$

which says that a belief J_1 is sublated at a certain time and for all time to follow that (t_{2+}) if and only there is a time prior to t_2 (here called t_1) and a person *x* such that *x* has the belief J_1 in the existence of a certain content *c* at t_1, and such that at t_2 and for all time to follow *x* does not have a belief in the existence of that content. This fits Dharmarāja's definition (DD25), which says that sublation occurs when an effect (*kārya*) ceases (*vināśa*) along with its material support (*upādāna*). Here J_1, the belief, is the "effect" and its material support is that ignorance (*avidyā*) which is responsible for that belief; both of them cease for good at time t_{2+}.

The materials we have available for further construction are now increased to include the predicate "is sublated by," i.e., "B." In terms of "B" we can now derive Dharmarāja's second definition, the definition of true belief or knowledge, namely

DD2: Pramā (inclusive of memory) = df. abādhitārtha-viṣayakajñāna,[10]

as

$$\text{PD2: } Pr_j = \text{df. } (t)\ (\sim Bjt).$$

PD2 says that a belief J is true (Pr) if and only if *j* is never sublated, and that is precisely what Dharmarāja's second definition says.

The early part of Dharmarāja's work concerns the nature of perception. Let us try to reconstruct what he says on this score. First, we need a definition of a concept, that of the internal organ (*antaḥkaraṇa*), a concept he merely takes for granted but which we can't take for granted, given our method. Thus my PD3 doesn't have a corresponding definition in Dharmarāja's set of definitions.

$$\text{PD3: } Ant_{xjt} = \text{df. } (c)(t)(A_{jct} \supset Q_{xct})$$

PD3 says that an internal organ (*x*) is that thing whose limitation constitutes a belief, that is to say, that a belief is a limitation of an internal organ. An internal organ is conceived in Advaita as a kind of fluid substance which fills up the content of an awareness and takes on its form. When this happens Advaitins speak of an "operation" of the internal organ, and that is what PD4 defines.

$$\text{PD4: } AV_{xjt} = \text{df. } (y)(Ant_{yjt} \equiv Q_{yxt}).$$

---

10 VP, p. 3.

PD4 says that x is the operation of an internal organ relative to a certain belief j at a certain time t if and only if y is the internal organ responsible for j at t just providing that x, the operation, limits that internal organ at that time.

With these notions of an internal organ and its operation in hand, we are now ready to reconstruct Dharmarāja's seventh definition,

DD7: Pramāṇacaitanya = df. antaḥkaraṇavṛttyavacchinna-caitanya.[11]

This definition explicates the important notion of a *pramana* or instrument of true belief. Dharmarāja explains it as any belief that is limited by an operation of the internal organ. This is what my PD5 says also:

PD5: Pramāṇa_{jt} = df. $(\exists x)(AV_{xjt})$.

Dharmarāja's eighth definition

DD8: Pramātṛcaitanya = df. antaḥkaraṇāvacchinnacaitanya[12]

explains the correlative notion of a *pramātṛ*, i.e, the knower, the seat of true belief, and this is reconstructed in my

PD6: Pramātṛ_{jt} = df. $(\exists x)(\text{Ant}_{xjt}) \cdot (\exists c)(Q_{jct})$.

Now true beliefs may be mediate or immediate. Dharmarāja defines a perceptual belief as one that is immediate, or, as he puts it, in which there is no difference between the awareness that is the instrument of true belief and the awareness limited by the belief's content. (In an inference, a mediate awareness, there *is* such a difference since the awareness which is the true inferential belief is a different mental act from the awareness by which we immediately grasp the thing to be inferred.) This notion of Dharmarāja's is enunciated in his

DD4: Jñānagatapratyakṣa = df. pramāṇacaitanyasya viṣayāvacchinnacaitanyābheda,[13]

which is also the force of my

PD7: Pratyakṣa_{jct} = df. $\text{Pramāṇa}_{jt} \cdot Q_{jct}$.

In the *Vedāntaparibhāṣā* Dharmarāja proceeds from a definition of the

[11] VP, p. 12.
[12] VP, *ibid.*
[13] VP, *ibid.*

nature of a perceptual awareness to a definition (his fifth definition) of what it is for a content to be perceptible. As you can see from the list of his definitions, he first offers a relatively simple definition

DD5a: Viṣayagatapratyakṣa = df. (ghaṭāder) viṣayasya pramātṛabhinna.[14]

and then, after having defined what a *pramātṛ* is, he produces a more complicated and improved definition

DD5b: Viṣayagatapratyakṣa = df. svākāravṛttyupahita-pramātṛcaitanyasattātiriktasattāśūnyatve sati yogya[15]

which makes use of a notion of relative grades of *sattā* or existence. I want now to develop this latter notion, of grades of existence, in order to arrive at the reconstruction of Dharmarāja's improved D5b.

Broadly speaking, the Advaita notion of grades of existence involves a comparative notion of "greater reality than," which I try to capture in PD8.

PD8: $\text{Realer}_{xy} = \text{df. } (\exists z)(t)(L_{yzt} \cdot \sim L_{yxt} \cdot L_{xyt} \cdot \sim L_{zyt}$

This says that if two things are such that the second is located in (at, on) the first and not vice-versa, and there is some third thing such that the third thing is located in the second thing but not vice versa, then the first thing has greater existence—is "realer"—than the second thing. To take an example, Brahman has greater existence than the rope which we erroneously think to be a snake, since (according to Advaita theory) the rope is located in Brahman and the snake resides in the rope but not vice-versa.

PD9 and PD10 utilize this notion of grades of existence to introduce into the system the notions of the Unreal and the Real respectively. Something completely Unreal (*asat*), such as a hare's horn, or the son of a barren woman, is explained as that which has nothing that has either greater or lesser existence than it.

PD9: $\text{Unreal}_x = \text{df. } \sim(\exists y)(\text{Realer}_{yx}) \cdot \sim(\exists y)(\text{Realer}_{xy})$

This is because we are understanding *sat*, (*relative*) reality or existence, to belong to everything other than *asat*, which has no existence at all and thus cannot have more or less existence than something else. I capture this idea by appealing to the understanding of the primitive term "located," which must be taken to apply to only those things which are loci or are located

14 VP, p. 20.
15 VP, p. 25.

at, in or on loci. *Asat*, the Unreal, is neither a locus nor located.

On the other hand, for something to be Real, that is, to have the highest grade of existence, is to be such that it has a higher grade of reality than something else and to be such that nothing has a higher grade of reality than it does.

PD10: $\text{Real}_x = \text{df. } (\exists y)(\text{Realer}_{xy}) \cdot \sim(\exists z)(\text{Realer}_{zx})$

Advaitins, of course, believe there is only *one* such thing, and that that is Brahman, the true Self. So we have now defined Brahman into the system.

Everything which is neither ultimately Real nor completely Unreal is termed *anirvacanīya* in Advaita. PD11 formulates that idea.

PD11: $\text{Anirvacanīya}_x = \text{df. } \sim(x \text{ is Real}) \cdot \sim(x \text{ is Unreal})$.

We are working toward a reconstruction of Dharmarāja's improved fifth definition. That definition contains one more notion that has not yet been defined. This is the notion of being *upahita*, i.e., of being qualified by an *upādhi* or merely accidental qualifier. Being *upahita* or accidentally qualified is to be contrasted with being *viśiṣṭa* or essentially qualified, i.e., with being qualified by a resident qualifier (*viśeṣaṇa*) which, because necessarily residing there as part of the thing's nature, does of course qualify the thing in question. Dharmarāja's own definitions, DD14 and DD15

DD14: Viśe*ṣaṇa* = df. kāryānvayi vyāvartaka (vartamāna)[16]
DD15: Upādhi = df. kāryānanvayi vyāvartaka vartamāna[17]

are stated very succinctly. A *viśeṣaṇa* or resident qualifier, he says, is one that demarcates a thing while residing in it, while an *upādhi* or accidental qualifier demarcates a thing while not residing in it. My PD12 and PD13 simply formulate these two definitions in my systematic reconstruction.

PD12: $\text{Viśiṣṭa}_{xyt} = \text{df. } L_{xyt} \cdot Q_{xyt}$
PD13: $\text{Upahita}_{xyt} = \text{df} \cdot \sim L_{xyt} \cdot Q_{xyt}$

Now we are ready to work on Dharmarāja's DD5b. It says that a content *c* is perceptible if (1) it is the kind of thing fit (*yogya*) to be perceived, and (2) it lacks any greater existence than the existence possessed by the subject (the *pramātṛ* of the perceptual awareness) apparently qualified by the form of *c*.

PD14: $\text{Perceptible}_{cjt}\, x \supset \text{Up}_{jct} x \cdot \sim \text{Realer}_{cj}$

16 VP, p. 30.
17 VP, ibid.

PD14 reconstructs not the complete definition of perceptibility but only its second clause. The first clause of Dharmarāja's DD5b requires that the content in question be fit to be perceived, i.e., that it be the type of thing graspable by the relevant sense-organ, that it be present and not past or future, etc. These requirements are not general ones and will require very complicated predicates to allow their formulation, since we must introduce definitions of the various sense-realms in order to distinguish a visible perceptibilium from an audible one, etc. Goodman's system suggests how this might be done, but I shall not attempt it here.

Dharmarāja's twelfth definition illustrates how he uses epistemic concepts to derive ontological ones. The notion of an individual self or *jīva* is a fundamental one for Advaita. It can be defined given the vocabulary Dharmarāja has provided for us to this point. The definition is simple: an individual self is the consciousness which is actually qualified by "its" internal organ.

DD12: Jīva = df. antaḥkaraṇāvacchinnacaitanya[18]

("*Avacchinna*" here must be understood to have the sense of "*viśiṣṭa*" in order to distinguish DD12 from DD13, below.) If we think of consciousness as a sort of liquid material which flows into and fills up whatever channels and containers are present for it, then we can think of the internal organ as what limits consciousness, a "portion" of consciousness undergoes operations through a set of channels or tubes—the sense-organs and the mind—which direct consciousness in characteristic ways. This is not my analogy; it is Dharmarāja's himself. I quote a passage from the *Vedānta-paribhāṣā:*

> Here, just as the water of a tank, going out through a hole, and entering fields through channels, comes to have, even like those (fields), a quadrangular or other figure, similarly, the internal organ too, which is of the nature of light, going out through the sense of sight, etc. and reaching to the locality of contents like pot, is modified in the form of contents like pot. This. . . is called a psychosis (*vṛtti*).[19] (S.S. Suryanarayana Sastry translation, Adyar Library 1942, p. 13)

Dharmarāja says that consciousness is residently qualified by its internal organ since the internal organ is a constitutive feature of the individual consciousness, that is, the *jīva*. I reconstruct this as:

PD15: Jīva_{jt} = df. $(\exists x)(\text{Ant}_{xjt} \cdot \text{Viśiṣṭa}_{jxt})$

[18] VP, p. 29.
[19] VP, p. 13.

He distinguishes the individual consciousness, the *jīva*, from what he calls the *jīvasākṣin*, the individual witness, which he defines as the consciousness which is only accidentally qualified by its internal organ.

DD13: Jīvasākṣin = df. antaḥkaraṇopahitacaitanya[20]
PD16: Jīvasākṣin_{jt} = df. $(\exists x)(\text{Ant}_{xjt} \cdot \text{Upahita}_{jxt})$

The distinction between the *jīva* and the *jīvasākṣin* is this: The internal organ is, as Dharmarāja says, "of the nature of light," but it is not itself the source of that light. Rather, it is by nature unconscious, *jaḍa*, but reflects the light of pure consciousness. Being itself unconscious, the internal organ is not capable in itself being *about* anything. So the relation between consciousness and an internal organ is just that the internal organ residently qualifies consciousness, limiting it in such a way that it *constitutes* what we call that individual self. By contrast, when we speak of an individual self's being aware of something or other, we are identifying that self not through its constitutive character—that particular internal organ—but rather through its experiences, the contents it "witnesses," and since these contents are actually the contents of consciousness *per se*, and not the contents of the internal organ which differentiates this self, the witnessing self is being identified through an *upādhi*, through an accidental rather than a resident feature of it.

I have spent some time reconstructing Dharmarāja's earliest definitions, including most of his first 15 along with the 25th, the definition of sublation. In the *Vedāntaparibhāṣā* Dharmarāja goes on to define many more Advaita concepts, including God, the critical relations of transformation (*pariṇāma*) and manifestation (*vivarta*), inference, memory, the nature of the pervasion that links the probans and the probandum in inference, the nature of falsity, various critical notions that play a part in the theory of sentence-meaning, the remaining *pramāṇas* such as comparison, presumption and nonperception, truth (*prāmāṇya*) and intrinsic truth (*svataḥprāmāṇya*), the two kinds of defining features of Brahman, the period between creative ages (*pralaya*), the stages of awareness such as waking, sleep and deep sleep, various moral qualities, and liberation itself. My hypothesis is that one can follow through the entire set of 84 definitions and reconstruct them along the lines I have illustrated, using just these atoms and primitives or perhaps a set of atoms and primitives even more economical than these.

There is one additional aspect which arises from this way of reconstructing Advaita, one which may have occurred to you as I developed by reconstruction. A Western reader confronting Advaita, especially post-

---

[20] VP, p. 30.

Śaṃkara Advaita, for the first time and without the assistance of an imaginative teacher will, I think inevitably gain an impression of a theory which is almost entirely alien to his own common sense. This may or may not appeal to him—many Westerners come to Indian philosophy searching for something alien which may *replace* Western common sense, with which they feel fed up or alienated from already. But whether it comes as a surprise or a relief, the concepts of Advaita as they are likely to come to the unprepared reader will sound very foreign. And as a result it will be difficult for him to be able to distinguish an understanding of Advaita that is confused from one that is accurate. If Advaita categories and concepts are *that* foreign to Western common sense, he may reflect, then what appears to me to be their apparent inadequacy, inaccuracy and incoherence is merely a result of his own inability to get into tune with Indian ways. Unfortunately, the result of this reflection may be that he becomes entirely uncritical about Advaita as a philosophical account, tolerating all sorts of philosophical abuses on the ground that he is in no position to criticize them. This, I think we must agree, would be unfortunate, especially if, as I think is the case, Advaita is put forward as a viewpoint which does in principle satisfy the criteria of adequacy, accuracy and consistency, etc., of which we have spoken.

It is very relevant to this point, then, that the systematic reconstruction of Advaita Vedānta—or at any rate, Dharmarāja's version of it—provides as atoms and primitive predicates terms which are easily understandable by the Westerner. To put it another way, by deriving Advaita notions from atoms and predicates that the Western reader can understand nonmetaphorically, he is provided with the tools to distinguish between what the Advaitin intends literally and what he offers only figuratively. A recurrent difficulty for a non-Indian reading translation of Indian philosophical texts is that he is regularly misled by connotations of the English terms used to translate Sanskrit technical terms, connotations which the Indian translator is only dimly aware of and which were in no way on the mind of the original author whose work is being translated. But by reconstructing Advaita along the lines I have tried to illustrate, these connotations and metaphorical extensions are minimized, since every technical term is derived eventually from those few primitive notions that provide the basis for the reconstruction. To give examples: a term like "sublation," which is the common translation of the Sanskrit "*bādha*," may appear queer and foreign to a contemporary English reader, for that word is now archaic in his language. But the definition provided by Dharmarāja and reconstructed by me explicates it as applying to any belief which was held but has come to be disbelieved, a familiar notion for which we happen to have no easy current term but which we can

easily understand and apply. Likewise, say, with the notion of an "internal organ" (*antaḥkaraṇa*), a translation which may mean a variety of things to a variety of readers, but the definition of which in my reconstruction of Dharmarāja becomes intelligible and reasonably unambiguous.

This kind of clarification through systematization seems a necessary step on the way to a fair assessment of Advaita. The sample suggests the importance of attempting such a reconstruction even though one knows full well he lacks the means or the will to complete the task. As the reconstruction proceeds, along the lines I have proposed, the student will begin to formulate an assessment of the philosophy being reconstructed and whether it is worthwhile to continue the process. After all, a philosophy is something one wants for some purpose or other (or perhaps, for all purposes); once one becomes convinced that a given type of philosophy is helpful or unhelpful for one's purposes he will likely stop reconstructing and either get on with his purposes or try another system, as the case may be.

Though there may be one or more understandings in which Advaita becomes a mystical philosophy, that does not mean that it cannot be developed in a rigorous, systematic fashion. Indeed, that was done, most noticeably by Dharmarāja. The brief examination we have paid to the system that results suggests that Advaita categories are surprisingly capable of representing a great deal about what we need to understand in order to retain (or gain) a conviction that liberation is available to one who wishes to obtain it. Whether it is the best, or the worst, of the several Indian philosophical systems—each of which attempts to serve the same purpose of grounding that conviction—depends on whether as one proceeds in the systematization one finds oneself unable to continue without flouting one or more of the fundamental criteria of success, adequacy, accuracy, etc. It has less to do with one's predilections for synthesis or analysis, for mystical or mysterious concepts as opposed to rational and rigorous ones, or any such antecedent prejudices. One may, to be sure, find or fail to find Advaita's language attractive, but that is superficial. The proof of its worth comes when one understands it in one's own ordinary language and tests it against relevant criteria in comparison with other candidates that propose to serve the same purposes. Though there is something to the notion that different cultures and linguistic communities have different conceptual schemes, that does not imply that the schemes are entirely incommensurate, does not imply radical cultural relativism. In this way we may hope to understand why Advaita, despite its apparently alien terminology and its uncompromising monism and spiritualism, has appealed so widely to intellectuals in India, and appears to do so still.

This page intentionally left blank

KISOR CHAKRABARTI

THE NYĀYA-VAIŚEṢIKA THEORY OF UNIVERSALS

I. STATEMENT OF THE THEORY

The Nyāya theory of universals is unambiguously realistic and its realism is evident from the way a universal (*sāmānya* or *jāti*) is defined. A universal is defined as "an entity that is eternal and inseparably inherent in many entities".[1] Since a universal is an eternal entity that is present in many particulars, it is clear that a universal is conceived to have an ontological status that is distinct from and independent of that of the particulars themselves. Thus in the Nyāya view there is an identical cloth-ness in all pieces of cloth, an identical humanity in all human beings; and the entitative status of these universals is not affected by the origin and destruction of particulars to which they may be related.

The relation between a universal and its particulars (called *samavāya,* which for want of anything better, we have translated as the relation of 'inseparable inherence') is held to be of a unique kind and sharply distinguished from ordinary relations like that of contact (*saṃyoga*) between two substances, relations of magnitude (*parimāṇa*), spatial and temporal relations, etc. This relation is held to be an 'inseparable relation' (*ayutasiddhavṛtti*) in the sense that one of the two relata, viz., the particular, has necessarily to remain related to the other relatum, viz., the universal, until it is destroyed. The necessity and inseparability of the relation is, therefore, only in one direction and not in both directions; and in fact it has been specifically mentioned that a universal would persist, even if all objects belonging to the class happened to be destroyed.[2] There is no reason to confuse this relation with an internal relation as has been done by modern interpreters[3], because the necessity of the relation is only in one direction and not in both directions.

As for the evidence in justification of the above theory, one common argument is derived from 'the notion of sharing a common character' (*anugatapratīti*). Thus in the Nyāya view, though individual human beings differ from one another in innumerable respects, we discern a common element of humanity running through all of them, though it may not be always possible to specify in words what that common element is. It is argued that in the absence of any compelling evidence to the contrary, such knowledge of identity (which presumably is present in every human mind) has to be

regarded as true, whence the admission of universals as real entities becomes necessary as the objective basis of such knowledge of identity. On the ground of this knowledge of identity, we classify objects (in a broad sense) into different natural kinds, e.g., man, horse, colour, sound, etc. Without the admission of universals, such classification has to be regarded as ultimately arbitrary and conventional.

A second argument concerns the significance of general words in our language. A general word like 'horse' or 'colour' refers to an unlimited number of objects and an explanation is needed as to how a word is capable of doing so. An explanation may be found if universals are admitted to be objectively real. A distinction has to be made between the *śakya* or denotation and *śakyatāvacchedaka* or connotation[4] of a word. Then it may be said that a general word connotes a universal and it is able to denote an unlimited number of objects, because the universal is present in all of them. Without admitting universals it would be difficult to explain how the relationship between a word and the unlimited number of objects it may denote happens to be established.[5]

A third argument concerns the justification of causal laws. The statement of a 'causal law' (*sāmānyakāraṇatā*) implies that every object of a certain kind (and nothing else) is the cause of a particular kind of effect. Thus when we determine fire to be the cause of smoke what we mean is that all fires (of a certain kind) are causes of smoke and that nothing but fire is the cause of smoke. As the Nyāya would say, 'the property of being the cause of smoke' (*dhūmakāraṇatā*) is possessed by all fires and by nothing else. But how should we explain (from the philosophical point of view) the fact that all objects of a certain kind become the cause, and nothing else becomes the cause? According to the Nyāya, the only way to explain this rationally is to suppose that all those objects share a common characteristic by virtue of which they can become the cause, and since this characteristic is not possessed by other kinds of objects, they do not become the cause. The admission of such common characteristics shared by many objects, however, amounts to the admission of universals.[6]

A fourth argument, closely related to the previous one, concerns the justification of 'inductive generalisation' (*vyāptigraha*). A causal law is a general statement which involves a generalisation from the observation of a limited number of particular instances. The problem is to justify the 'inductive leap' from the observation of a limited number of cases to the

generalisation which covers not only cases observed, but also all possible cases, past, present and future. It is argued that a solution to the 'problem of induction' is found by admitting universals as real entities. Then it may be said that from an observation of particular instances we are able to see the connection between the universal properties involved; and the generalisation to all cases would be justified, because identical properties are present in all cases.[7]

On the basis of the arguments (rather superficially) outlined above, the Nyāya considers it necessary to admit the existence of universals as objectively real entities which are distinct from, and independent of, particulars themselves. The range of universals admitted to be objectively real is rather comprehensive. There are first of all what may be called generic universals, i.e., universals whose instances are individual substances (*dravya*), for example, man, horse or stone. There are, secondly, the qualitative universals, i.e., universals whose instances are qualities (*guṇa*), for example, colour, sound or shape. The notion of quality (*guṇa*) is very broad in Nyāya metaphysics. Thus in addition to ordinary qualities of physical objects, there are qualities of the 'soul', such as pleasure, pain, desiring, hating etc., and there are universals corresponding to these qualities as well as to physical qualities. Moreover, many relations, such as relations of magnitude (*parimāṇa*), the relation of contact (*saṃyoga*) between two substances, e.g., the book is *on* the table, relations of spatial or temporal proximity or distance etc., are brought under the head of qualities; and universals pertaining to them are also admitted to exist. Thirdly, there are universals pertaining to motion (*karma*), such as contraction, expansion, etc. Russell[8] has noted that though philosophers have recognised universals named by adjectives and substantives, those named by verbs and prepositions have been usually overlooked. A look at the above list will show that this complaint certainly could not be brought against the Nyāya theory.

With respect to extension, universals are classified into three kinds.[9] There is first the universal of largest extension (*parasāmānya*), viz., existence (*sattā*) in which every other universal is included. There are, secondly, universals of smallest extension (*aparasāmānya*), e.g., cowness, horseness etc. in which no other universal is included. Most universals, however, are such that while they are included in universals of larger extension, they themselves include universals of smaller extension within them. For example, the class of animals is included in the class of substances and itself includes classes like man, horse etc. within it. This classification corresponds exactly to the Aristotelian

view that except for the highest genus (*summum genus*) and the lowest species (*infimae species*), whatever is a species is at the same time a genus and whatever is a genus is also a species.

Further light may be thrown on the Nyāya theory by comparing it with similar theories in Western philosophy. The Nyāya theory should be clearly distinguished from the *Ante Rem* or Ideal theory of the Platonic type. According to the latter theory, Ideas or Forms which are the reality behind natural objects, exist eternally in the world of Being and have their shadowy manifestations in the world of Becoming. We know the ultimate Forms by *a priori* intuition and so are able to recognise their incomplete manifestations in sensible things. In the Nyāya view, however, particulars are as much real as universals themselves, and the relation between the former and the latter is that of 'inseparable inherence'. On this point the Nyāya theory is much closer to the Aristotelian theory. Thus, like Aristotle, the Nyāya would deny that there is (as stated in the tenth book of Plato's Republic) an unseen and eternal bed, or quintessential bedness, by participation in which any bed becomes a bed; but still would hold, like Aristotle, that there is an identical bedness in all beds, an identical humanity in Smith, Brown and Jones, and the relation between such identities and particulars to which they belong is not merely accidental, but necessary (in the sense specified above).[10]

Moreover, Aristotle distinguishes between primary and secondary beings (substances) and holds that though an individual man or horse is a primary being, the species man or horse is a secondary being. Thus the universal, for Aristotle, could not exist as an individual, a primary being, exists. The individual cannot be repeated, and cannot be common to many. Consequently, that which is universal and common to many cannot *be* in the sense in which the individual is. Turning now to the Nyāya theory we find that a distinction is made between *sattā* and *bhāvatva*, which may be translated as 'existence' and 'being' respectively. It is held that though universals have *bhāvatva* or being, they do not have *satta* or existence; while particulars have both 'being' and 'existence'. The distinction between 'existence' and 'being' is not at all clear and would require careful research to be spelled out exactly. But one thing that is clear is that though universals and particulars are both real, Nyāya is making a distinction between the ontological status of the two, a distinction which is a permanent part of Aristotle's teaching. It may also be noted that Russell[11] distinguished between 'existence' and 'subsistence' and held that while particulars exist, universals subsist or have being, where

subsistence is opposed to existence as being timeless.

For Aristotle, however, a universal is an abstraction and therefore non-sensible. Both it and its relations are apprehended by reason. (In Plato's view also Ideas or Forms are non-sensible and can be grasped only by *a priori* intuition.) But in the Nyāya view most universals (with the exception of universals like 'atomness', 'soulness' etc.) are known through sense perception. Thus in the Nyāya view, when three different pots are presented together in perception, we may observe that all of them share a common character, viz., pot-ness. The Nyāya does not deny that dispositional elements are present, both in the recognition of the common character and in the observation that two or more objects have the same character. But in either case, none of this makes the discovery less of a discovery. Again, in the sense of singling one element out of what is given, abstraction is necessarily present in the discovery of a common character like colour. We concentrate upon the colour in noticing that the colour in three different objects is one and the same, and single it out for observation. But such observation does not involve abstraction in another sense, viz., framing an abstract complex idea. In the Nyāya view a universal is not an abstraction in the latter sense, and this is true not only for qualitative universals like colour, sound, etc,, but also for generic universals like horseness and cowness. Thus cowness is a simple, indivisible element that is present in every cow and is directly apprehended at the moment an individual cow is perceived. The Nyāya view on this point is quite uncompromising; the reason behind this view is that predicates such as 'cow', 'horse', etc. are regarded as simple and unanalysable, as are qualitative predicates like 'red', 'blue', etc. In fact in the Nyāya view it is only a simple predicate that could name a universal. If a predicate were complex and made out of simpler concept, it could not name a universal because the universal is a unitary entity devoid of any parts. Thus cowness is not a complex idea formed by putting together the common features of different individual cows. On this point the Nyāya view is diametrically opposed to the Lockeian view that general ideas such as 'cow', 'man', etc., are framed through a process of conceptual combination. As already said, if a general idea is formed through a process of conceptual combination, then in the Nyāya view, it cannot be the name of a universal. On the other hand cowness, humanity, etc. are simple properties and discovered directly through sense perception.

This also shows how different the Nyāya view is from the Aristotelian view regarding the relation between a genus and its species. In Aristotle's

view the essence of the species man, for example, is constituted by the genus animality together with the specific difference rationality. Thus in Aristotle's view the relation between a genus and its species is necessary, the genus constituting a part of the essence of the species. In the Nyāya view, however, both humanity and animality are simple concepts. The relation between the two is not necessary in the Aristotelian sense, because animality is not a part of the concept of man.

This difference has resulted in a basic difference between the theories of definition of Nyāya and Aristotle. In Aristotle's view a definition must be the statement of essence, the statement *per genus et differentiam*. In the Nyāya view, such essential definition of terms like cow and man is impossible. On the other hand, according to the Nyāya, a definition is the statement of a 'unique characteristic' (*asādhāraṇadharma*), i.e., a characteristic which is present in every member of the class of objects to be defined and not present in anything else.[12] The genus is never included in a definition since a genus is not a unique characteristic in the above sense. Moreover, such a unique characteristic does not have to be essential in Aristotle's sense, but may very well be accidental. For example, cow is defined as that which has a dewlap. Now dewlap is a unique characteristic of cows in the sense specified, and hence this definition is satisfactory from the Nyāya point of view. But a dewlap is still an 'accident' and hence for Aristotle this definition is not satisfactory. In Aristotle's view the aim of a definition is to state what the definiendum really is. In the Nyāya view the aim of definition is much more modest. The aim is to differentiate the definiendum from everything else and be able to use words unambiguously. Both of these two aims are fulfilled by the statement of a unique characteristic.

Another peculiarity of the Nyāya theory must now be stated. To a Western realist a quality is a repeatable character, such as the colour blue. Nyāya, however, distinguishes between particular qualities (*guṇa*) which are not repeatable and common qualities which are repeatable. Thus each blue substance is connected with its own blue colour which is not shared at all and is as particular as the substance itself; all these distinct blue colours are connected in turn to the shared universal blueness. Just as there is a class of men so there is a class of 'blues' and the members of this class are not blue substances, but the distinct blue colours themselves. A similar distinction is made for relations. There are relation-particulars as well as relation-universals; and relation-universals have as their members the relation-particulars and not the relata

themselves. It may be noted that whether qualities and relations of particular things are particular or universal is a controversial question in contemporary philosophy.[13] The Nyāya position in this controversy would be that in one sense all qualities and relations of particular things are also particular. But particular qualities and relations may themselves be classified under common qualities and relations and these latter are all universal.

II. OBJECTIONS TO THE THEORY BY OTHER SCHOOLS

The Nyāya theory of universals came under serious criticism from many other schools of Indian philosophy; criticisms often followed the pattern of criticisms of realistic theories in Western philosophy. The Nyāya argued that different individuals are known to be identical in certain respects and universals must be admitted as the basis for this knowledge of identity. The Sāṃkhya challenged what is being assumed in this argument, viz., that different individuals are known to be identical in certain respects. It was objected that different individuals are not known to be identical but only known to be similar in some respects. Three different pots may be known to be similar to one another because all of them have a similar colour, say, red. And the colour in respect of which these are similar is also not the identical colour in all of them. The colour of the first pot is as particular as the pot itself and is only similar to the colour of the second pot. At this point it was held by Śālikanātha[14] of the Mīmāṃsā school that what different individuals really have in common is similarity or resemblance. Hence it is enough to admit resemblance (*sādṛśya*) itself as a distinct entity and thus get rid of the innumerable universals hypostatised in the Nyāya theory.

The controversy of similarity vs identity was carried on indefinitely and we will only briefly indicate how the Nyāya defended its position. It was argued that mere similarity is not enough because "individuals of different kinds are also known to be similar".[15] The point here is that 'belonging to the same class' or 'being of the same kind' is a much stricter notion than being merely similar. For example, a chimpanzee may be said to be similar to an orangutang, but it would be improper to say that the same relationship holds between a chimpanzee and another chimpanzee. It is also unwarranted to regard similarity as a distinct entity, because similarity between two entities always reduces to their having certain characteristics in common.[16] When, however, we come to consider the common characteristics themselves, it goes

contrary to experience to insist that these characteristics are merely known to be similar and not identical. Let us consider two pots having exactly the same shade of red. In such a case the redness in the first instance of red is experientially indistinguishable from that of the second instance of red. One may say that the first red is distinguishable from the second red, because they are after all in different places. But that is completely beside the point. Nyāya grants that as an instance of red the first red is different from the second. What is indistinguishable are not the instances of red, but what they are instances of. When we consider what they are instances of, we have to ask whether they are instances of the same shade or not? If they are not, the point is irrelevant. If they are, then what makes them instances of the same shade must be the presence of that shade in each. To insist on mere similarity with respect to the shade is to insist on a difference that is not discernible and, therefore, has no empirical evidence to back it up.

The most persistent critics of the Nyāya theory were the Buddhists who pointed to a series of difficulties in the Nyāya theory. Granted that the universal is a distinct entity, it has to be asked whether it is present in a particular in its entirety or only in part. Neither alternative is acceptable. If the universal is present in its entirety in one of the particulars, it could not be present in other particulars. Nor can we say that it exists in a particular only in part, because it, ex hypothesi, is simple and partless. It may be noted that this objection is basically the same as that raised by Parmenides against the doctrine of Forms in the dialogue *Parmenides* of Plato.[17]

Again, if the universal is distinct from particulars, it becomes necessary to explain how the universal comes to be related to a new-born particular, e.g., a cow. We cannot say that the universal has moved from the place where it already existed to the place where the cow is born, because a universal is not a substance; and according to the Nyāya, only substances are capable of motion. Nor can we say that cowness already existed at the place where the cow was born, because then it should have been perceived there even before the cow was born. It is also not possible for us to say that cowness also originated at the same place along with the cow, because, by hypothesis, it is eternal. In a word, the admission of universals as distinct entities raises difficulties about the relationship between the universal and particulars.[18]

Moreover, the Nyāya argued that universals must be admitted on the basis of the notion of belonging to the same class. It was pointed out that this argument is inconclusive, because there are cases where we do have the

notion of belonging to the same class but no common simple character may be found to be possessed by all members of the class. For example, let us consider the class of cooks. As Ashoka Pandita said: "Though with reference to many individual cooks, there is a common notion of all of them being a cook, there is nothing which may be common to all of them."[19] Ashoka Pandita disposes of the suggestion that what individual cooks have in common is the act of cooking and shows by detailed analysis that there is no common simple character possessed by all cooks. Whatever may be the merit of the actual example chosen, the important point scored by the Buddhist is that if we could speak of the class of cooks without requiring a common character, why could we not do so with respect to other classes like man, horse etc.?

The Buddhists held that all that exist are particulars and that universals are nothing but 'conceptual constructions' (*vikalpa*). A radically different explanation of the significance of general words was also given. When we apply a general word, such as *cow,* to an individual cow, we do not mean that it is of the same kind as other cows. What we really mean is that it is different from everything that is other than a cow. By applying the word *cow,* we rule out the application of other words such as *horse, man,* etc. and differentiate the object from every non-cow. Thus a general word primarily has a negative meaning signifying 'differentiation from others' (*anyāpoha*) and not a positive meaning signifying 'belonging to the same class' as in the Nyāya view.[20]

The Nyāya replied that before we can differentiate cow from every non-cow, we must already know what is meant by *cow*. The concept of cow is part of the concept of differentiation from every non-cow and is logically presupposed by the latter. The Buddhist analysis of the meaning of general words does not succeed in dispensing with the positive meaning of them, but rather confirms it. In fact the analysis is vitiated by circularity. As Jayanta remarks, "If the meaning of cow is to be ascertained through the negation of non-cow, circularity is inevitable. 'Cow' is to be ascertained through negation of non-cow, but negation of non-cow is possible only through an ascertainment of what cow is."[21]

As to the difficulties raised, the Nyāya position is that they are due to a misunderstanding of the nature of universals. When it is asked whether a universal resides entirely or partially in a particular, it is presupposed that a universal must reside either entirely or partially in a particular. But "Cowness is neither a substantial whole (*avayavī*), nor an aggregate; the word 'part' applies to members of an aggregate or to elements of a substantial whole; the

word 'entire' applies to such members or elements when all of them are taken together without a remainder. Cowness is neither an aggregate nor a substantial whole; hence the words 'entire' and 'partial' are not applicable to it."[22] Thus by a subtle analysis the objection is removed by showing that the alternatives contemplated are not applicable to universals at all.

The other difficulty is also due to not realising the basic difference between a universal and a particular. A particular cannot exist at more than one place at the same time; but a universal, by hypothesis, is capable of residing at many places at the same time. So the natural thing to say is that a universal 'resides in all objects belonging to the class' (*svaviṣayasarvagata*). So the universal also resides in a new member that happens to be added to the class by being produced and there is nothing mysterious or problematic about it. When an object is produced, 'the sum total of causal conditions' (*kāraṇasāmagrī*) determines its nature and thus to which class it should belong or what kind it should be. So if somebody chooses to wonder how the relationship between a universal and a new member of the class is established, the answer is that it is the causes producing the object that establish the relationship between it and the universal.[23]

As to the remaining objection the Nyāya firmly held to the position that the admission of universals for man, cow, etc., was justified on the basis of experience. Thus at the time of perceiving two individual men, we directly observe that both share the common property of humanity and this kind of direct experience cannot be nullified by an appeal to cases like the class of cooks. The Nyāya, however, conceded the point that there are cases where in spite of the 'notion of belonging to the same class', no universals could be admitted. We must distinguish between general terms such as *man, cow*, etc. which name universals and other general terms such as *cook, father*, etc. which do not name any universals. In fact the Nyāya held that before a universal is admitted on the basis of the knowledge of identity, we must ascertain that no violation has been made of any of the 'restictive conditions for universals' (*jātibādhaka*) to the discussion of which we now turn in the following section.

III. RESTRICTIVE CONDITIONS FOR UNIVERSALS

Granted that there are universals, an important problem for the realist is to determine whether there are universals answering to every common name, and if not, how the population of the world of universals can be limited to

what is reasonable. This problem has figured prominently both in Plato's philosophy and in Nyāya. We may first briefly indicate Plato's position regarding this problem.

It is asserted in the *Republic*[24] that there is an Idea corresponding to every common name, but there are substantial grounds for doubting whether Plato ever seriously upheld this doctrine. There is first the question whether in Plato's view there are ideas corresponding to negative terms such as not-man, not-beautiful etc. Plato's view on this point is not altogether free from ambiguity. In the *Sophist* he seems to be saying that there is an Idea corresponding to not-beautiful as much as there is one corresponding to beautiful. But in the *Politics* he says in a definitive way that though terms like not-Greek, 'not-ten-thousand' stand for parts of the genera man and number, they do not stand for species of them, implying that there is no Idea of not-Greek or of not-ten-thousand. Secondly, in the *Parmenides,* upon being asked by Parmenides Socrates says that though he recognises Ideas of goodness, beauty, justice, etc., he is more hesitant about the recognition of Ideas of man, fire, water, etc. and that he definitely does not admit ideas of mud, dirt and hair. This is followed by Parmenides' remark that Socrates' refusal to recognise the latter Ideas is due to philosophical immaturity and that he should recognise them following the general principle. This exchange of opinion between Parmenides and Socrates indicates that though Plato found it necessary to admit ideas for ethical and aesthetic values, he was probably more doubtful about extending the theory of Ideas to such 'unpleasant' and 'trivial' things as mud, dirt, etc. It seems that the basic reason for the theory of Ideas was to provide an objective foundation for art and morality as well as an objective foundation for knowledge (which prompted admission of Ideas for mathematical objects). From this point of view Plato was doubtless less concerned with things like mud and dirt, because they are, after all, not objects of knowledge in the stricter Platonic sense of knowledge that is necessarily true. Thirdly, it becomes clear from the remarks of Aristotle that some Platonists held that there are Ideas only of substances and that there are no Ideas of such artificially produced things as house, ring, etc.

All this evidence suggests that according to Plato and some Platonists there is not a universal answering to every common name – a view very clearly and unambiguously upheld by the Nyāya. One basic restrictive condition as already mentioned is that a universal must be simple, i.e., it must not be analysable into other properties or property components. For

this reason a general term like *cow* or *man* would stand for a universal, but not terms like *white man* or *black cow*. The latter terms represent 'complex properties' (*sakhaṇḍopādhi*) and do not name additional, ontologically distinct entities. That is to say, the property of being a black cow is not an additional entity over and above blackness and cowness, but is reducible to them. For the same reason negative terms such as not-man, not-red, etc., do not stand for universals. Any negation presupposes an idea of what is negated and hence cannot be regarded as simple. Moreover, many terms such as deafness, blindness, etc. which do not have the form of a negative term, but are negative in meaning, also do not stand for universals. (These terms are comparable to privative terms of Aristotelian logic.) Similarly, relative terms like father, brother, etc. cannot be said to stand for universals, because from the very nature of the case they cannot be regarded as simple.[25]

A second restrictive condition is that whenever a universal is admitted, the relationship between it and its particulars must be 'necessary' or 'inseparable' in the required sense. For example, let us consider the class of cooks mentioned earlier. Though we can speak of a class of cooks, 'cookness' or the property of being a cook is not a universal because the relationship between it and the individuals is not necessary, but accidental.

A third restrictive condition is as follows. Two universals may be so related that the first is included in the second, i.e., every member of the first is a member of the second, e.g., humanity and animality. Again, two universals may be so related that neither is included in the other, e.g., cowness and horseness. It is held that in the latter case two such universals could not have any members in common; in other words, there must not be any cross-division or over-lapping of universals. So two universals may have some members in common if and only if the first is included in the second or vice versa.[26]

A fourth restrictive condition is that no universal can be admitted where the admission would result in violation of the essential nature of members. Thus no universal can be admitted to be shared in common by the so-called 'ultimate individuators' (*viśesa*) which are held to be self-differentiating and which are hypostatised to differentiate qualitatively indistinguishable atoms from one another. (Atoms cannot be held to be different from one another without such hypostatised individuators, because it would have followed by an application of the principle of identity of indiscernibles that atoms which are qualitatively indistinguishable are all one and the same.) Objects of the

same kind can always be differentiated from objects of other kinds through their difference in kind. Thus a cow can be differentiated from a horse through the universal cowness possessed by the former and not possessed by the latter. If all ultimate individuators are held to be of one kind sharing a common universal, we would also have to hold that they could be differentiated from other entities through the universal possessed by them. But that would be contrary to the essential nature of ultimate individuators, viz., that they are self-differentiating. (It is also necessary to hold that they are self-differentiating. If we ask how an ultimate individuator can be differentiated from another ultimate individuator and postulate that they are distinguished from each other through another ultimate individuator, we would be hurried off into a vicious infinite regress.) Hence the admission of no such universal can be allowed.

According to a fifth restrictive condition, if two universals have exactly the same members, they are not two different universals, but one. For example, potness and 'conch-shell-like-neck-ness' (*kambugrīvādimatva*) are instantiated in exactly the same entities, viz., pots, and hence there are not two different universals corresponding to them, but one. It is obvious that this restrictive condition states what is basically the same as the thesis of extensionality for sets in modern set theory. This restrictive condition was introduced by Udayana, the last great logician belonging to the school of early Nyāya, and in all probablity he is the first logician to have formulated the extensionality thesis[27] This restrictive condition, however, reflects an important change in the concept of a universal. While in the early Nyāya view a universal is understood as a class-property or a repeatable character, because of this restrictive condition a universal (*jāti*), for Udayana, comes to be regarded as a class in extension. In fact Udayana held that as an 'abstract property' (*upādhi*), 'conch-shell-like-neck-ness' is different from potness. This shows that Udayana distinguished between a universal (*jāti*) or a class in extension and an abstract property (*upādhi*), a distinction which is never made in earlier Nyāya philosophy. It appears that in Udayana's view corresponding to the same class in extension, there may be more than one abstract property, a view which is also upheld by many modern logicians.[28] In fact, in the process of discussing the restrictive conditions, Udayana threw new light on the problem of universals and opened up a whole new perspective as may be seen from the account of the following restrictive conditions.

A sixth restrictive condition is that no universal can be admitted to exist,

the admission of which would lead to a vicious infinite regress. For this reason there can be no universal of which every universal is a member; for if we had any such universal, then, by hypothesis, we have a given totality of all universals that exist and all of them belong to this big universal. But this universal is itself a universal and hence (since it cannot be a member of itself, because in Udayana's view no universal can be a member of itself) it too along with others must belong to a bigger universal and so on ad infinitum. What is said here has interesting analogues in modern set theory in which it is held that a set of all sets (i.e., a set to which every set belongs) does not exist. And the reason why the set of all sets cannot exist is similar to that given here, viz., that if such a set existed, it must itself belong to a still larger set and this process will be continued ad infinitum; so that in a certain sense it must be larger than itself, which is a contradiction.[29]

Udayana of course has no objection to infinite regress as such, but does object to infinite regress of a certain kind which he considers to be vicious. An infinite regress is vicious if it makes the 'basis of the regress' (*mūla*) impossible. Thus the infinite regress arising in the case of the universal of all universals is vicious, because what the regress essentially proves is that we could never have such a universal of all universals. There are other cases of infinite regress which are not vicious in the above sense. For example, if one asks whether the tree is the cause of the seed or the seed is the cause of the tree, one would find oneself thrown into a regress that is unstoppable. This regress is not vicious, because in this case the regress starts only after supposing that the tree or seed already exists. The situation is different with the universal of all universals. Here we could not suppose that it already exists because the very concept of a universal of all universals commits us to an infinite regress. In fact if we could suppose that the universal of all universals already exists, the regress would not have started at all.[30]

A seventh restrictive condition is that no universal can be admitted where the relation of 'inseparable inherence' (*samavāya*) between the putative universal and its members could not be admitted to be possible. We know that the relation between a universal and its members is called the relation of 'inseparable inherence'. We could ask whether there is a universal which has inseparable inherence as a member, and the answer is said to be no. The reason why the answer is no is that the relation of inseparable inherence could not be admitted to be possible between the universal and inseparable inherence itself. For suppose that the relation of inseparable inherence is possible in a

given case. Then we have to suppose that inseparable inherence is related to the universal by inseparable inherence. But the latter is also a member of the universal and hence it also must be related to it by inseparable inherence and so on to a vicious infinite regress. The vicious infinite regress makes the admission of the relation of inseparable inherence impossible in the given case and that in turn rules out the admission of the universal. It is probable that there are cases where the relation of inseparable inherence would be impossible for reasons other than vicious infinite regress and probably that is why this restrictive condition has been stated separately as an additional restrictive condition and not a sub case of the previous one.

On this point too we can see an interesting parallel in modern set theory. The relation between a set and its elements is the relation of membership. As it was asked whether there is a universal whose member is inseparable inherence, so it may be asked whether there is a set whose member is membership and the answer is no. This set may be expressed in set-theoretical terms as the set of all ordered pairs $\langle x, y \rangle$ such that $x \epsilon y$, which may be written with the standard notation as follows: $\{\langle x, y \rangle \mid x \epsilon y\}$. Assuming that this set exists, we can talk about its domain D: $D\ \{\langle x, y \rangle \mid x \epsilon y\} = \{x \mid \exists y\,(x \epsilon y)\}$. But given any set x, it is a member of some set, e.g., $\{x\}$ From this it follows that $\forall x \exists y\,(x \epsilon y)$; from which it follows that $\forall x\, x \epsilon D$.

Thus D is proved to be the universal set which as already seen in the discussion of the previous restrictive condition, could not exist; and hence it follows that the above set also could not exist.

Finally, a property that could not be instantiated in more than one object is not a universal. The very basis of admitting a universal, viz., the notion of belonging to the same class, is impossible in such a case. For example, *diktva* or spaceness is not a universal, because space is one and infinite, and hence the property spaceness is instantiated only in one object. Spaceness, however, is a 'simple property' (*akhaṇḍopādhi*) and like a universal is held to a real entity not ontologically distinct from its locus, viz., space.

By implication a property which could be instantiated in no objects at all is also not a universal; and precisely for the same reason as above, viz., that the notion of belonging to the same class is impossible in such a case. For example, general terms such as 'rabbit's horn', 'son of a barren mother', 'turtle's hair', etc., do not represent universals. Udayana devoted considerable attention to the analysis of such 'non-referring expressions' and at the risk of a slight digression we may briefly state the gist of it. We may begin with an

examination of statements like 'the rabbit's horn does not exist'; in nine out of ten cases the subject term of a sentence denotes an object. But to hold that the subject term of our sentence denotes an object is clearly unsatisfactory, because the very meaning of the sentence is that there is no such object as a rabbit's horn. Again let us consider the statement 'the rabbit's horn is not sharp'. It may seem that the statement is true; since the rabbit's horn does not exist, it is false to say that it is sharp and hence the statement that it is not sharp ought to be true. But Udayana points out that both the statements 'the rabbit's horn is sharp' and 'the rabbit's horn is not sharp' are false and proper analysis will show that there is nothing paradoxical about it. The statement 'the rabbit's horn is sharp' is much more complex than it may apparently seem and ought to be analysed as follows: 'the rabbit exists, the horn exists, the horn is sharp and the rabbit possesses the horn'. Now it is clear that the statement is false, because one of the conjuncts, viz., that 'the rabbit possesses the horn' is false. For the same reason the statement 'the rabbit's horn is not sharp' is also false. Similarly, the statement 'the rabbit's horn does not exist' should be analysed as 'no rabbit has a horn' in which case the apparent difficulty of not finding any denotation for the subject term is removed.[31]

It should be noted that this analysis of 'non-denoting expressions' is remarkably similar to that of Russell.[32] Russell also held that both the statements 'the present king of France is bald' and 'the present king of France is not-bald' are false and gave a similar analysis of these two statements. Russell rejected[33] Frege's view that definite descriptions always have denotation on the ground that this creates problems for expressions like 'the present king of France' where the denotation appears to be absent. Nyāya, likewise, would not' subscribe to the position that all definite descriptions are denoting expressions. In the Nyāya terminology a definite description like 'the son of Yajñadatta' stands for a 'complex predicate' (*sakhaṇḍopādhi*). The expression of a complex predicate is a denoting expression when every component of it is the name of a real entity. Thus in our example 'the son of Yajñadatta' is a denoting expression, because every component, viz., 'son', 'Yajñadatta' and the relation 'of', is the name of a real entity. But 'the rabbit's horn' is not a denoting expression, because one of the components, viz., possession of horn by rabbit, is not the name of a real entity.

We may now bring together the results we have so far obtained. There are first of all terms like man, cow, etc., corresponding to which there are universals which are ontologically real entities distinct from individual

objects. Secondly,there are terms like spaceness, etc., corresponding to which there are no universals, but are called 'simple properties' which are also ontologically real and distinct from their loci. Thirdly, there are terms like 'black cow', 'white man', etc., which represent complex properties and do not stand for any new, additional universals over and above those represented by their simple components. The peculiarity of the Nyāya view is that both indefinite descriptions like 'a Brahmin from Kashi' and definite descriptions like 'the son of Yajñadatta' are thrown into the class of complex properties. Fourthly, there are terms like 'rabbit's horn', 'son of a barren mother', etc., which, unlike those listed above, are 'non-denoting expressions'. Finally, there are terms like 'the universal of all universals', 'the universal whose member is inseparable inherence', etc., which could not represent any universals because of the special difficulties involved.[34]

To conclude: We have seen that the Nyāya and Vaiśeṣika believed that universals are real entities ontologically distinct from particulars and inseparably inherent in them. The distinctive feature of the theory is that universals are held to be simple as well as sensible. Another distinctive feature is an elaborate study of various restrictive conditions that prevent the admission of universals answering to every common name. Our discussion shows that though one may not necessarily agree with either the main tenet of the theory or specific conclusions drawn on particular points, as a theory it is worked out extremely well to minute detail requiring subtle philosophical analysis. For all these reasons the Nyāya-Vaiśeṣika theory should be regarded as an interesting counterpart of the theories of universals in Western philosophy.

University of Calcutta

NOTES

[1] See (20), pp. 51 – 2. The number in parentheses refers to the number of the entry in the bibliography. Though the theory of universals presented in the paper was developed by the Vaiśeṣika as much as by the Nyāya school, we have avoided writing down the word Vaiśeṣika every time a reference to the Nyāya view has been made.

[2] See (20), p. 59.

[3] See (8), section on *samavāya*.

[4] We have used the word connotation for the want of anything better. It should here be understood to mean 'the mode under which an entity becomes presented through

the word' and is comparable to Frege's *Sinn.* For Frege's view, see 'On Sense and Nominatum' reprinted in (2).

[5] The first two arguments have been concisely stated by Śrīdhara as follows: "... the nature of each object being different from others, how could there be knowledge of a common form in them (without the universal), and how could they be designated by the same word? (Without the universal) the relation of a word with the unlimited individual objects of a class could not be established." See (17), p. 32.

[6] See (15), pp. 80 ff.

[7] See (15), pp. 458 ff.

[8] See (12), p. 94.

[9] The classification is also often said to be twofold, viz., (a) the universal of largest extension and (b) universals with an extension smaller than that of another universal. Raghunātha of Navya-Nyāya, however, held that there is no universal of the largest extension like *sattā* or 'existence' in which every universal is included. See (20), pp. 59–60.

[10] Radhakrishnan writes: "The universals on this view, answer to the separate, supra-sensual arch-typal forms. ... Praśastapada's view is akin to Plato's realism, according to which sensible things are what they are by participation in the universal forms of Ideas which are eternal and self-subsistent." See (10), vol. 2, pp. 211–2. Our remarks show that such assimilation of the Nyāya view to Plato's view is a mistake.

[11] See (12), P. 100.

[12] See (20), p. 52. A definition must be non-circular and free from the defects of being too wide, too narrow and impossible. (Impossibility here means not applying to any definiendum at all.) A definition must also be 'simple' and avoid superflous qualification.

[13] See (1), chap. IX.

[14] See (13), chap. on Pramāṇapārāyaṇa.

[15] "Mere similarity is not enough as held by Sāṃkhya, because individuals of different kinds are also known to be similar. ..." See (7), p. 204.

[16] See (20), p. 27.

[17] See (3), vol . VI, p. 215.

[18] See (17), p. 755. For a discussion of this and other Buddhist objections also see (15), chap. IX.

[19] See (16), p. 94.

[20] On this point the Buddhist view is comparable to Spinoza's view that all determination is negation. For a fuller account of the Buddhist view see (14), Introduction.

[21] See (4), p. 278.

[22] See (18), p. 669.

[23] See (17), p. 754.

[24] For the necessary references in this paragraph see (11), chap. XI.

[25] See (5), p. 73.

[26] For this and following restrictive conditions see (4), pp. 73–7. This particular restrictive condition and the next restrictive condition were, however, rejected by Raghunātha see (20), p. 56.

[27] Udayana's date is held to be 10th or 11th century A.D. The author is not aware of any earlier logician who could be claimed to have formulated the extensionality thesis. For Udayana's date see (10), vol. 2, pp. 152–3.

[28] See (9), p. 88.

[29] See (6), p. 170.

[30] The distinction between an infinite regress that is vicious and an infinite regress that is not vicious is a very important logico-mathematical discovery and the author proposes to unravel the full significance of this distinction in a future paper.

[31] See (5), pp. 133–40.

[32] See Russell's 'On Denoting' reprinted in (2), pp. 108, 112.

[33] See (2), p. 107.

[34] A very significant development that took place in post-Udayana Nyāya philosophy (usually known as Navyanyāya) was the generalisation of the concept of *upādhi* already introduced by Udayana. Navyanyāya logicians used the term *upādhi* to mean any class of objects, irrespective of the number of objects in the class, irrespective of whether objects in the class shared a simple, common property, irrespective of whether the relation between the objects and the class was 'inseparable'. Everything that exists came to be regarded as an *upādhi* in this generalised sense. It appears that the Navyanyāya concept of *upādhi* was a close approximation to the modern concept of a set. A full discussion of the Navyanyāya theory of *upādhi* is, however, beyond the scope of this paper and has to be left to be the subject-matter of a series of papers in future.

BIBLIOGRAPHY

Aaron, R. I., *The Theory of Universals* (Oxford: Clarendon Press, 1952).

Feigl, H. and Sellars, eds., *Readings in Philosophical Analysis* (Appleton-Century-Crofts, 1949).

Fowler, H. N., ed., *Plato* (William Heinemann, 1917).

Jayanta, *Nyāyamañjarī* (Benares, Chowkhamba: Kashi Sanskrit Series No. 106, 1936).

Maitra, S. K., *Fundamental Questions of Metaphysics and Logic*, Calcutta, 1956.

Matilal, B. K., *Epistemology, Logic and Grammar in Indian Philosophical Analysis* (The Hague, Mouton, 1971).

Mendelson, E., *Introduction to Mathematical Logic* (D. Van Nostrand, 1964).

Mishra, U., *History of Indian Philosophy*, Vol. II (Allahabad, Tirabhukti Publications, 1966).

Pap, A., *An Introduction to Philosophy of Science* (New York, Free Press, 1962).

Radhakrishnan, S., *Indian Philosophy*, Vols. I and II, (London, George Allen and Unwin, 1929).

Ross, D., *Plato's Theory of Ideas* (Oxford, Clarendon Press, 1963).

Russell, B., *The Problems of Philosophy* (Oxford University Press, 1912).

Śālikanātha, *Prakaraṇapañcikā* (Benares, Hindu University, 1961).

Sharma, D., *The Differentiation Theory of Meaning in Indian Logic* (The Hague, Mouton, 1969).

Shastri, D. N., *Critique of Indian Realism* (Agra University, Agra, 1964).

Shastri, H., ed., *Six Buddhist Nyāya Tracts* (Calcutta, Bibliotheca Indica, 1910).

Śrīdhara, *Nyāyakandalī* and *Padārthadharmasaṃgraha* of Praśastapāda, ed. with a Hindi translation and introduction by Jha, D. (Benares Sanskrit University, 1963).

Tarkatirtha, A. and Tarkatirtha, T., eds., *Nyāyadarśana* with *Bhāsya, Vārttika, Tātparyaṭīkā* and *Vṛtti* (Calcutta Sanskrit Series, 1936).

Udayana, *Kiraṇāvali*, eds. Sarvabhauma, S. and Vedantatirtha, N. C. (Calcutta, Bibliotheca Indica, 1956).
Viśvanātha, *Bhāṣāpariccheda* and *Siddhāntamuktāvalī*, ed. with a Bengali translation by Shastri, P. (Calcutta, Sanskrit Pustak Bhandar, 3rd. ed., 1969).

MARK SIDERITS

REVIEW ARTICLE

MORE THINGS IN HEAVEN AND EARTH

In his work, *The Problem of Universals in Indian Philosophy,*[1] Raja Ram Dravid has done an excellent job of locating and elucidating a large number of the major discussions of the problem of universals in the Indian tradition. In so doing, he succeeds in demonstrating that all of the major positions which are associated with the problem in the Western tradition – extreme realism, moderate realism, conceptualism, and extreme nominalism – are to be found as well in Indian philosophy. The picture which emerges from his account is that of an often heated debate in which the strengths and weaknesses of various approaches to the problem are thoroughly explored. Dravid's discussions of this debate reveal him to be possessed of a remarkable command of an extensive and widely varied literature. He has done other scholars in the field of Indian philosophy a great service by providing an illuminating guide to the relevant portions of a sometimes bewildering array of texts.

There are, nonetheless, a number of problematic areas in Dravid's work. In what follows I shall be focusing on these, not because I wish to belittle Dravid's accomplishment, but rather because I feel there are some important methodological issues here which require comment. The general question which informs my remarks is this: To what use should we put those materials on the problem of universals which we discover in the Indian tradition. This question arises in part because it is not clear that Dravid feels there is much, if anything, which we can learn about the problem from the examination of these materials. I wish to suggest that a study of the Indian debate on the question of universals can play an important role in our attempts at understanding this issue, for instance by suggesting to us possible approaches which have not been extensively explored in the Western tradition. I shall confine my remarks to two issues: the status of qualities, and the possibility of a consistent extreme nominalism. On the first point, the consensus in at least the modern Western tradition is that if there are universals, these are of the nature of recurrent qualities. In the Indian tradition, on the other hand, it is assumed by realists and nominalists alike that the qualities of a particular

substance are themselves particular. If there are universals, these form a category distinct from the category of qualities. Dravid takes exception to this presupposition, and I shall comment on the grounds of his rejection. Again, it is generally accepted in the modern Western tradition that a consistent radical nominalism is unattainable. Thus most modern forms of nominalism are variations on the theme of the resemblance theory. Within the Indian tradition, on the other hand, we find one school, the Yogācāra-Sautrāntika, devoting an immense amount of effort to the task of constructing an extreme nominalism which makes do without the relation of resemblance. Dravid finds the resulting theory, apohavāda, extremely implausible. I shall try to show that this theory is at least deserving of much closer scrutiny, and that it has not yet been demonstrated that the project fails.

Dravid points out that within Indian philosophy the possibility is never considered that there might be universal qualities or relations. In the Nyāya-Vaiśeṣika categorial scheme, qualities and relations have an ontological status different from that of universals, and qualities and relations are assumed to be particulars. While universals may inhere in individual instances of red or contact (though not in the relation of inherence), redness and contactness are not to be thought of as themselves qualities or relations. It would seem that this doctrine set the tone for subsequent discussions of the problem of universals, all of which focus on the question, whether there are real entities which are common to certain classes of particular substances, qualities, and relations. Dravid clearly indicates that he finds this feature of the Indian examination defective. His own position is that we directly experience identical, recurrent qualities, relations, and resemblances. It is on the basis of such experience that we construct such general concepts as the concept of cow or the concept of red. Such distinct class essences as 'cowness' or 'redness' are, then, wholly superfluous (p. 347).

It is thus interesting to note that, in his discussion of some contemporary views on the problem of universals, Dravid takes strong exception to Stout's doctrine of the abstract particular. Of the latter's contention that the properties of a concrete particular are themselves particular, Dravid says, 'If it were true, then we could not assert of two things A and B that both are "red" or both are "round" ' (p. 433). No evidence is given for this somewhat startling claim, and this is not surprising, for Dravid has already described two answers to the question how, if qualities are particulars, we could say of A and B that both are red: (1) Stout's answer, that there is a 'distributive unity' which

determines the class of abstract particulars which are called 'red'; and (2) the Nyāya answer, that we perceive the universal redness in both the particular color of A and the particular color of B.

Dravid also accuses Stout of dogmatism on the point of the latter's insistence that the characteristics of a concrete particular are themselves particular. It is far from clear, however, that Dravid is not equally dogmatic in his own, oft-repeated, insistence that experience presents us with countless instances of identical, recurrent qualities. He correctly points out that Stout's chief argument in defense of the theory of abstract particulars is inconclusive. This argument is that since any two concrete particulars A and B must themselves be locally separate, it follows that the properties of A and the properties of B must likewise be locally separate; from this it is said to follow that any property of A must be distinct from any property of B, since two entities can be said to be identical only if they have the same spatial location. Of course the conclusion does not follow, for while spatial location is among the criteria of identity for physical objects, it may well be illegitimate to employ this criterion in determining the identity of properties of physical objects.

Dravid's own position is on equally shaky footing, however. I take him to be saying that when, under ideal observation conditions, we are prepared to affirm that two objects A and B are the same shade of red, we must conclude that these qualities of the two objects are strictly identical. It is instructive to consider this situation from the perspective of radical translation, for then we shall see that Quine's indeterminacy thesis applies here as well. Suppose that we demonstrate to the speaker A and B, and receive an affirmative response to the query which we have established to correspond to the English, 'The same red?' Does this suffice to establish that the speaker accepts an ontology of recurrent qualities? There are any number of other possible accounts of what the speaker has assented to: 'A and B more nearly resemble the tail of a gavagai than they do any other particular;' 'The (particular) quality inhering in A and the (particular) quality inhering in B are both inhered in by redness'; and so on. Nor will more complete knowledge of the syntactical devices whereby individuation, demonstration and the like are achieved in the speaker's language suffice definitively to rule out any of these many translational variants. With sufficient ingenuity we can still attribute to the speakers of that language an ontology of resembling particulars or of particular qualities and quality-universals, for accommodating devices can always be

incorporated into other portions of our translation manual. And all of this applies, with equal force, to the home language as well. In general, for this as for other parts of ontology, experience underdetermines theory, and syntax and semantics provide nothing more than suggestions. When we employ the English locution, 'the same shade of red,' are we committed to the existence of a single quality which occurs in distinct particulars? Once again, it seems as if only a lack of ingenuity at providing alternative accounts of the structure of this expression leads us to suppose that we are so committed.

Perhaps Dravid's case is not quite as bad as I have made it out to be. He does, after all, provide some argumentation for his general position on the problem of universals, and it is possible that an argument could be constructed, from these materials, in support of his stand against quality particulars. In this case we would no longer be forced to view him as relying exclusively on appeals to experience and our linguistic intuitions. Thus he argues that extreme realism involves insuperable difficulties on such points as the relation of universal to particular (p. 458). He claims that extreme nominalism leaves unexplained our ability to use general terms in a non-arbitrary fashion (p. 459). Finally, he objects to an Aristotelian type of moderate realism, which accepts an ontology of subsistent class-essences (humanity, etc.), by bringing up the Lockean appeal to borderline cases (p. 460). One might then argue that we can account for our ability to employ general terms only by appealing to recurrent qualities and relations.

It is noteworthy, however, that among the relations which Dravid allows into his ontology is resemblance. This is done presumably to forestall the objection that many of the qualitative similarities which we experience are not instances of exact qualitative identity. Such a position is inherently unstable, for on the one hand a Lockean objection, to the effect that the existence of borderline cases undercuts talk of qualitative identity, might be pushed here as well; and on the other hand, it might be urged that the concept of resemblance is derivative and eliminable in favor of qualitative identity. For instance, the existence of borderline cases might be taken to show that the divisions between those qualities which constitute the range of shades of red are so fine-grained as to be beyond the capacities of our perceptual apparatus to notice. In this case, what we take to be a single shade of red is in fact a set of discrete shades. This explains not only the existence of borderline cases but also our use of the notion of resemblance. We construct the pseudo-quality red out of discrete shade of red precisely

because we take the occurrence of borderline cases as evidence for the existence of a continuum; and we then posit the relation of resemblance in respect of redness to account for the relations among the points of this apparent continuum.

Thus such a line of argument would seem to establish not Dravid's conclusion, that there are recurrent qualities and relations, but rather that our ability to employ general terms must be based either on real resemblances or on qualitative identities. Furthermore, it is far from clear that the denial of quality particulars follows from either of these alternatives. A resemblance theory can just as readily be constructed on the basis of an ontology which includes particular qualities as on the basis of an ontology consisting only of particular substances. One might favor the substance-based ontology on grounds of economy, but if we allow particular qualities into our ontology as well, this will at least have the result of simplifying the task of constructing resemblance classes. Suppose we choose, instead of the resemblance theory, to allow talk of qualitative identity. In this case the option is still open to us of treating qualities as individuated by the substances in which they inhere (that is, treating quality-terms as degenerate mass-nouns), and explaining talk of qualitative identity by appealing to talk of co-inherence of quality-universals. Here again the principle of parsimony might be invoked against such a supposition, but it is possible that epistemological riches can be bought at the price of a little ontological heaviness. The notion that we can perceive a qualitative identity between two particulars is problematic, for we can always ask what it is about the constitution of A and the constitution of B which leads us to judge them as sharing an identical quality. When we ask this, we seem to be inquiring about the particular causal capacities of A and the particular causal capacities of B; it is no longer clear that we must be talking of an identical quality. If we take as primitive the notions of quality particular and quality universal, we may well be able to sort this all out.

Finally, Dravid would seem to be somewhat over-hasty in his rejection of extreme and moderate realism, and extreme nominalism. For instance, his rejection of moderate realism is based on Lockean considerations about the difficulties we encounter when we seek to classify borderline cases by means of natural kind terms. Presumably, this shows that the middle-sized objects of our acquaintance do not fall naturally into determinate classes, and thus that our classification of such objects into natural kinds is in some sense arbitrary. Hence, it is argued, these objects do not possess real essences. There

is, however, a Kripkean objection to this line of argument, to the effect that our ability to use natural kind terms in modal contexts shows that such terms must be construed as rigid designators: The statement, 'Water is H_2O', while synthetic, is necessary, that is, true in all possible worlds. If this is correct, then the door is once again open to a doctrine of real essences for natural kinds. And once we allow substance universals into our ontology, there remains no obvious bar to treating the qualities inhering in substances in a similar fashion, namely as particulars sharing a common essence. I am not myself sure of the strength of this Kripkean line of argument. It does seem clear, however, that if (as seems to be the case) Dravid's stand against quality particulars depends on his rejection of essentialism along with such other theories as extreme realism and extreme nominalism, then his own stance is not sufficiently defended by the Lockean borderline cases argument alone; further argumentation is required to establish the falsity of essentialism.

Dravid's rejection of the extreme realism of Nyāya-Vaiśeṣika and the extreme nominalism of Yogācāra-Sautrāntika seems likewise over-hasty. Here it would seem that extra caution is called for, as there is more at stake. We have been considering an argument which Dravid might use in support of his claim that qualities are not particulars, an argument in which the rejection of extreme and moderate realism and extreme nominalism plays a role. The question of the ontological status of qualities is but a part of the problem of universals, and perhaps one on which not a great deal hangs. We want to know as well whether there is good reason to abandon those theories which Dravid rejects, and it is possible that a close examination of the Indian materials will help us here. Now as it turns out, there is little on the question of essences in the Indian tradition that strikes us as novel. There is, however, a great deal that strikes us as new in the treatment of extreme realism and extreme nominalism. What I wish to suggest is that Dravid has missed the novelty here, and thus raises objections to the Indian versions of these theories which are inappropriate in that context. In fact, I suspect that we can learn a great deal about the strengths and weaknesses of extreme nominalism and extreme realism by considering them in their Indian formulations, for then we may see that certain of the stock Western objections apply only to special cases of these theories. In what follows I shall attempt to show that Dravid's objections to the extreme nominalism of *apohavāda*, at any rate, misfire. In so doing, I hope to show how an examination of the Indian controversy may contribute to our understanding of the problem of universals.

Dravid begins his estimate of *apohavāda* by acknowledging the extent to which this theory is bound up with the metaphysics of Yogācāra-Sautrāntika; he appears to conclude from this that since the latter does not allow of direct disproof, the former can only be assessed in terms of relative adequacy to our experience. Immediately thereafter, however, he claims that in the absence of recurrent qualities and relations, conceptual cognition would be impossible (p. 340). This would seem to suggest that the metaphysical basis of apohavāda is not merely relatively inadequate to our experience, but quite strictly false to our experience. He then goes on to maintain that the principal issue at stake between the apohavādin and his opponent is whether the meaning of a word includes the negation of the opposite meaning. Dravid asserts both that the meaning of a positive word is felt as positive, and that the negation of the opposite meaning would be impossible if the word did not already possess a positive meaning (p. 342). He also finds the claim that absolutely unique particulars generate universal concepts 'quite unintelligible'. If the particulars are without any resemblances whatever, then it seems that the extension of a concept must necessarily be arbitrary (pp. 344–5).

There are, in addition to these, a number of other criticisms presented by Dravid, but these will suffice for our purposes. What I wish to suggest is that he would have done well to take more seriously some important differences between what is at least one of the major Indian conceptions of the philosophical enterprise and a certain modern Western conception of that enterprise. Dravid repeatedly insists that there are recurrent qualities and relations. On this point he probably has the support of common sense. We do, after all, speak of a curtain and a tablecloth as having the same color. The testimony of common sense is not, however, entirely unambiguous, on this or on other questions of ontology. Thus what we seem to find quite frequently in Indian philosophy are attempts at constructing internally consistent ontologies in relative isolation from common sense, followed by the development of what we might call bridge theories, which are meant to show how our precritical intuitions may be reconciled with the details of the favored ontology. In Yogācārara-Sautrāntika we find on ontology of unique, spatially and temporally atomic particulars, the *svalakṣaṇas*. I would suggest that we look upon *apohavāda* as a sort of bridge theory, one which is designed to account for the facts of linguistic practice on the basis of such an ontology. If this suggestion is correct, then it is wrong to criticize the Yogācāra-Sautrāntika ontology until the details of *apohavāda* (as well as other bridge theories such

as the theory of judgment) have been fully grasped. Nor is it at all germane to argue that the meaning of a positive word is felt as positive. Presumably, whether we feel the meaning of a word to be positive or negative will depend in large part on the theory we hold; and since the näive semantic realism of common sense would seem to suggest that the meaning of a positive word is positive, it is hardly surprising that this is our intuition. The relevant question to ask here is whether *apohavāda* can furnish an adequate explanation of this datum, given that there is no room for positive meanings in the *svalakṣaṇa* ontology. Again, it may well be our intuition that 'in the absence of a positive meaning the negation of the opposite meaning is inconceivable' (p. 342). It does not follow from this, however, that any attempt at accounting for this intuition which does not admit 'positive meaning' as a primitive notion is doomed to failure. Again it seems question-begging at best baldly to assert that two particulars can cause the occurrence of identical concepts only if the causal capacities of the two are identical. It is among the tasks of the apohavādin to show that we need invoke neither real universals nor resemblances among particulars to account for determinacy of extension in our concepts; but in order to understand his strategy here, we must first examine his analysis of generality, for that analysis plays a key part in his approach to the question of the causal capacity of particulars. This approach may ultimately prove inadequate, but it would seem incumbent on us to make greater efforts to grasp the underlying analysis before we decide that the theory has failed in this respect.

What I am suggesting, in short, is that we view *apohavāda* as part of a translation manual, rather than as an independent semantic theory. The picture I have in mind is this. On the one hand there is the set of common-sense beliefs about the world; on the other hand there is a set of beliefs about a world populated only by *svalakṣaṇas*. A fully successful translation manual would give a truth-preserving translation of each member of the former set into some member(s) of the latter set. Of course such a translation manual should be written in accordance with the usual constraints on theory construction: The application of translation rules should be straight-forward and unambiguous; rules which appear to be entirely *ad hoc* are not to be admitted; other things being equal, the fewer rules the better; etc. Thus if we are presented with two rival ontologies, both equipped with similarly adequate translation manuals (where the criterion of adequacy is ability to generate truth-preserving translations for a maximum number of common-

sense beliefs), we may judge between them on the basis of the extent to which their respective translation manuals embody these desiderata of theory-construction. On this view we are precluded, however, from taking any component of the translation manual in isolation and judging it on the basis of its adequacy to our experience.

I would claim that this is the chief problem with Dravid's evaluation of *apohavāda*. In criticizing certain fundamental theses of *apohavāda* on the basis of our intuitions about linguistic meaning, he suggests that he sees this theory as one which is meant to replace the common-sense view of language. I would suggest instead that the Yogācāra-Sautrāntika seeks to show that the common-sense view does not commit us to the existence of universals, i.e., that there is a plausible explanation of linguistic behavior which does not require a realistic ontology. What is called into question by his doctrine are not the facts themselves, but rather the use which certain philosophers make of them. What his doctrine is meant to show is simply that we can give an adequate explanation of our intuitions about language without engaging in ontological inflation: For every meaningful ordinary-language sentence, there is an apohist sentence of equivalent meaning but lacking commitment to the existence of abstract entities. Here the facts are the same, our intuitions go unchallenged; the only possible grounds for objection lie in the translation rules themselves.

My interpretation of the Yogācāra-Sautrāntika enterprise stands in need of defense. What I propose to do is offer an attempt at a rational reconstruction of some key elements of the doctrine of *apoha*, basing this reconstruction on the view sketched above of the underlying methodology of the school. What I hope to show is that viewing *apohavāda* in this way leads to an interpretation of the doctrine which is more plausible and less easily refuted than is Dravid's interpretation. Such an interpretation would, I believe, help to vindicate my view in that it would better accord with the principle of charity than does Dravid's treatment of the apohavāda materials. The Yogācāra-Sautrāntika did, after all, maintain this doctrine for some six and a half centuries, and in the face of objections quite similar to those which Dravid puts forth. While it may be possible that they were all quite uniformly dense, I think we would do well to explore other possibilities before we take this view of their philosophical abilities.

My reconstruction is based on the 'Śabdārthaparikṣā' chapter of *Tattvasaṃgraha* by Śāntarakṣita with the *Pañjikā* of Kamalaśīla.[2] This text

is important for its discussion of the distinction between two types of negation, *prasajya-pratiṣeda* and *paryudāsa*. This distinction seems to have originated with the grammarians, and was employed extensively by the Mimāṃsakas in their analyses of Vedic injuctions. A standard explication of the distinction which is frequently cited in the literature is the following:

> Where there is subordinateness of the positive, primariness in denial,
> That is *prasajya-pratiṣedha*, where the 'not' is attached to the verb.
> Where there is primariness of the positive, subordinateness in denial,
> That is to be known as *paryudāsa*, where the 'not' is [attached] to a subsequent term.

Prasajya-pratiṣedha we shall refer to as verbally bound negation; and since by a 'subsequent term' is meant a noun or adjective, and adjectives and nouns tend to be treated similarly in Sanskrit, we shall call paryudāsa nominally bound negation.[3] Consider these two sentences:

(1) An armadillo is not a rodent.
(2) An armadillo is unhurried.

In (1) the negation may be said to be verbally bound; here the emphasis is on denial, not on something 'positive'. That is, an utterance of (1) would ordinarily be taken to rule out the assertion that an armadillo is a rodent, but would not in general be taken to commit the speaker to the claim that the armadillo is some other sort of animal than a rodent. While such a claim might be among the implicatures that are generated by an assertion of (1) in certain contexts, it would be implausible to suppose this to be the principal intention of the speaker. By contrast in (2), which is an instance of nominally bound negation, the element of denial is subordinated to the element of commitment: An utterance of this sentence would more plausibly be taken as an assertion that the armadillo is the opposite of hurried, i.e., slow, than as a mere denial that the armadillo is swift, unaccompanied by commitment to any other characterization of its locomotion.

The general claim which is embodied in the grammarians' distinction between *prasajya-pratiṣedha* and *paryudāsa* is that any negation which has the syntactic feature of being verbally bound will have as its chief force a denial with at best a minimal level of commitment, and that any negation which has the syntactic feature of being nominally bound will have as its chief force

commitment to some characterization of the subject. We shall not discuss the merits of this claim for the interpretation of Sanskrit, let alone for English. We shall, however, note in passing that if something like this distinction does underlie the use of negation in either language, then we should expect to find sentences which appear to violate the standard rule of double negation. And it seems that we do, for consider the sentence,

(3) What he did was not impolite.

If we construe the negative prefix 'im-' as a mere stylistic variant of straightforward negation, then (3) should mean the same as,

(4) What he did was not not polite.

By double negation, (4) yields

(5) What he did was polite.

It is clear, however, that (3) and (5) do not have the same meaning. The distinction between nominally and verbally bound negation helps us here. If we view the negative prefix 'im-' as an instance of nominally bound negation, and the 'not' as verbally bound negation, we see that in (3) both kinds of negation are present. We then realize that while the characterization of an action as impolite involves commitment to the claim that the action was rude, the denial, by means of verbally bound negation, of that claim does not involve commitment to any other characterization of the action; in particular, it does not involve commitment to the claim that the action was polite. It is for this reason that (3) and (5) differ in meaning.

One interesting feature of this distinction is the following: While verbally bound negation is a sentential connective, nominally bound negation operates only on sentential elements. Thus in (1) the 'not' produces the denial of the assertion, 'The armadillo is a rodent'. In (2), on the other hand, the negative prefix 'un-' operates only on the predicate 'hurried', yielding a new predicate which is asserted of the armadillo. The Naiyāyikas, for reasons having to do with their doctrine of cognition,[4] seek to construe all negations as operating an sentential elements. Thus the entire thrust of the Nyāya doctrine of absence is to show that in the case of a cognition corresponding to a negative statement, the negation may be taken as a component of the content of the cognition, and so should not be taken as a contribution of the cognizer. Given this motivation, the doctrine of absence has the consequence that

prasajya-pratiṣedha negations take on a central feature of paryudāsa, namely commitment. Thus take the following case of *prasajya-pratiṣedha*:

(6) The pot is not an ashtray.

Here we seem to have a model case of denial without commitment to any positive characterization of the pot. On the Nyāya analysis, however, (6) becomes,

(7) The pot possesses a mutual absence, the counterpositive of which is ashtray.

Whereas (6) appeared to be the denial of an identity statement, (7) is clearly the attribution of a certain property to the pot. The possibility of so treating all negative statements suggests that not only nominally bound negation but verbally bound negation as well involves commitment to the existence of some property.

Śāntarakṣita and Kamalaśīla accept the Nyāya strategy, but not its consequences. They agree that negation should be construed as operating on sentential elements, not on sentences; in this respect the notion of *apoha* resembles that of an absence. They deny, however, that all negation involves commitment to the existence of some property; in this respect they maintain the distinction between *prasajya-pratiṣedha* and *paryudāsa*, though now both types are taken as having the same restrictions on scope.

According to Śāntarakṣita, there are altogether three types of apoha: paryudāsa apoha pertaining to mental images (*buddhi* or *pratibhāsa*), paryudāsa apoha pertaining to the *svalakṣaṇas*, and prasajya-pratiṣedha apoha (TS 1003). The first and second types both being paryudāsa, these involve the element of commitment: to say that a given entity excludes, in this fashion, other entities is to give a positive characterization of the entity; one is thereby committed to the existence of this exclusion as a property which takes the given entity as locus. In this respect, these two types of apoha resemble a Nyāya absence. The two kinds of paryudāsa apoha differ in two respects. The first and most obvious difference is that they operate on different sorts of entities, the first on *pratibhāsas*, the second on *svalakṣaṇas*. What is excluded by the paryudāsa apoha of a mental image are other pratibhāsas and not svalakṣaṇas. This type of apoha has as its domain all the pratibhāsas which might occur, as mental contents, to a subject. Apoha pertaining to the svalakṣaṇa, on the other hand, ranges over the domain of

the reals, the *svalakṣaṇas*, forming exclusion classes from these. The second difference has to do with the size of an exclusion-class formed by an apoha. The paryudāsa apoha of a given svalakṣaṇa is the exclusion of all other svalakṣaṇas; but the same does not hold true of apoha pertaining to pratibhāsa. The exclusion class formed by the apoha of a given pratibhāsa contains most but not all of the other mental images which might occur. Those *pratibhāsa* which are not excluded by this apoha constitute, together with the given pratibhāsa, a set of images which correspond to the extension of a general term.

The last point is explained by way of the following analogy. The three distinct plants *abhaya, dhātṛ*, and *harītaka* are all known to possess anti-febrile properties. Ingestion of any one of these three plants produces, as a result, the abatement of fever. It is important to note, however, that this fact does not warrant the conclusion that the three distinct plants all possess the same causal capacity. We are justified only in concluding that each plant has the capacity to cure specific instances of fever, but not that there is some one causal property shared in by all three. By the same token, let us suppose that there are three mutually distinct *svalakṣaṇas*, each of which causes the occurrence of a distinct *pratibhāsa*. Each of these pratibhāsas, in turn, is such as to exclude the occurrence of a range of other pratibhāsas. That is, each pratibhāsa comes equipped, as it were, with its own paryudāsa apoha whereby it naturally generates an exclusion-class. Let us further suppose that each of our three original pratibhāsas generates the same exclusion-class, i.e., excludes the same particular pratibhāsas. It is then legitimate to speak of these pratibhāsas as caused by svalakṣaṇas which form part of the extension of one general term, even though so far we have said nothing which would justify the conclusion that either the svalakṣaṇas or the pratibhāsas which are their effects possess a common essence (TS 1004–5, 922).

One further step is required before we come to the apohavādin account of linguistic meaning. A central feature of the paryudāsa style of apoha is ontological commitment. To say that a svalakṣaṇa excludes all other svalakṣaṇas, or that a pratibhāsa excludes certain other pratibhāsas, is to give a positive characterization of the svalakṣaṇa or pratibhāsa, it is to assert that the entity is the locus of a certain property. This feature is absent from the prasajya-pratiṣedha style of apoha. Here we have instead a kind of exclusion which is an absolute rejection or prohibition, something which can perhaps be likened to the move in chess which consists of sweeping all the

pieces from the board. In *tarka* (dialectics), *prasajya-pratiṣedha* represents the denial not only of the thesis which has been reduced to absurdity, but of the conceptual scheme which produced the thesis as well; it is a way of rejecting thesis without committing oneself to antithesis.[5] With prasajya-pratiṣedha apoha, exclusion does not involve commitment to any such property as the absence of the excluded. What results is rather the pure rejection of a proposed characterization. If such an apoha is to be thought of as at all an entity, it must be thought of as no more than a constructed fiction. As Śāntarakṣita puts it, such an apoha is neither positive nor negative, it is neither an existent (bhāva) nor an absence (abhāva). (TS 1188).

We may now proceed to the apohavādin account of linguistic meaning. Here we must bear in mind that for both Nyāya and Mīmāṃsā the meaning of a general term such as 'cow' will involve the notion of the universal, i.e., cowness. The Yogācāra-Sautrāntika, of course, denies the existence of universals. His ontology, moreover, does not allow him to explain univocity in terms of real resemblances among svalakṣaṇas, for the latter are said to be absolutely unique. Thus he proposes the following analysis: The meaning of 'cow' is not non-cow. The key to this analysis lies, I believe, in the fact that the mental content 'not non-cow' is constructed with the use of two different types of apoha: the prefix 'non-' representing a paryudāsa apoha on pratibhāsas, and the particle 'not' representing a prasajya-pratiṣedha apoha on the former apoha. Let us use the dash to stand for paryudāsa, and the tilde to stand for prasajya-pratiṣedha. We shall consider a world consisting of six entities, $s_1, s_2, \ldots, s_6$. Each of these entities is the cause of a mental image, p_n, s_1 being the cause of p_1, etc. Suppose, moreover, that the following is the case:

$$-p_1 = \{p_4, p_5, p_6\}$$
$$-p_2 = \{p_4, p_5, p_6\}$$
$$-p_3 = \{p_4, p_5, p_6\}$$

That is, p_1, p_2, and p_3 form the same exclusion-classes; the occurrence of any one of these pratibhāsas prevents the occurrence of p_4, p_5, and p_6. We shall assume that s_1, s_2, and s_3 form the extension of 'cow' (where cows are thought of as simples), and that the subject has learned, through presentations of s_1 and s_2, the convention governing the use of the word 'cow': that this word is properly applied just in case a pratibhāsa occurs which forms the exclusion-

class $\{p_4, p_5, p_6\}$. The subject is now presented with an occurrence of p_3, and responds with the statement, 'This is a cow'. How are we to explain this without invoking a universal as the meaning of 'cow'?

The suggestion is that we translate 'This is a cow' into apohist language as '$p_3 = \sim -p_3$'. Our apohist sentence consists of three elements, namely 'p_3', '=', and '$\sim -p_3$'. We shall comment on each in turn. The first presents no difficulties, since in this context 'this' clearly refers to the cause of p_3, and we need invoke no common property to account for the speaker's ability to identify p_3 as 'this'. The substitution of the identity sign for the copula signals a transformation of a subject-predicate statement into an identity statement. We shall defer discussion of this transformation for the moment. The crucial question is this: What are we to make of the expression, '$\sim -p_3$'? In particular, how is this meant to replace 'cow'? We know that $-p_3$ is just the exclusion class $\{p_4, p_5, p_6\}$. If the $\sim -$ function obeyed double negation, then substitution of the argument p_3 would yield the class $\{p_1, p_2, p_3\}$, as that portion of the domain which is not excluded by *apoha* on the exclusion class $\{p_4, p_5, p_6\}$. That is, $\sim -p_3$ would then be a real property which characterized the class $\{p_1, p_2, p_3\}$. We know, however, that the $\sim -$ function does not obey double negation. The prasajya-pratiṣedha apoha of the exclusion class $-p_3$ is an absolute rejection of the exclusion class as a characterization of p_3: p_3 is *not* to be thought of as belonging to that set of mental images which it by nature excludes. Now it happens to be the case that this exclusion class also fails as a characterization of p_1 and p_2. We might then think of $\sim -p_3$ as that which determines the class of cow-images, $\{p_1, p_2, p_3\}$. This is not to say, however, that there is some one thing which is common to p_1, p_2, and p_3; for there is no such thing as $\sim -p_3$. Here we must recall the central feature of prasajya-pratiṣedha apoha: denial without commitment to the existence of a property. In short, $\sim -p_3$ is the 'form' (*ākāra, ābhāsa*) of p_3, that in virtue of which p_3 is correctly said to be a cow; but $\sim -p_3$ is not a real. The expression '$\sim -p_n$' serves merely to identify that mental operation whereby the extension of a term is determined.

One may, of course, think of $\sim -p_n$ as the meaning of a term; but here we must bear in mind the fact that the question, 'What is the meaning of "cow"?' can be misleading, particularly if it causes us to suppose that there is some entity which answers to the description, 'the meaning of "cow"'. This question is answered satisfactorily by any theory which succeeds in explaining how we are able to apply the term 'cow' to just those things which are cows. The apohavādin purports to have done just this. If his account proves satisfactory, then what

he has shown is that we can dispense entirely with the notion of meanings as entities: 'Linguistic meaning' is a mere convenient fiction.

I claim that the apohist translation of 'This is a cow' is '$p_3 = \sim -p_3$'. How is it that a subject-predicate statement is transformed into an identity statement? The first point to be made in this regard is that, strictly speaking, this identity statement is not true. For any n, p_n is not identical with $\sim -p_n$. In the present context, p_3 is the pratibhāsa or mental representation of s_3. $\sim -p_3$, on the other hand, is not the mental representation of any entity. Indeed the whole point of the distinction between prasajya-pratiṣedha apoha and paryudāsa apoha is to ensure that $\sim -p_n$ not be the same as p_n. The $\sim -$ function does not obey the standard rule of double negation. Śāntarakṣita's remark that an anyāpoha is neither positive nor negative (TS 1188) must be taken in this light. A second point, however, is that $\sim -p_3$ resembles nothing so much as it resembles p_3. As Kamalaśīla puts it in commenting on TS 1005, the $\sim -$ function yields an anyāpoha which 'contains the image of the object by means of the object-formness, there is an awareness which has the appearance of the object by means of identity'. Again, Śāntarakṣita states that this is called 'anyāpoha' 'because of its being considered, by those who are confused about it by reason of identity, in the form, what is the svalakṣaṇa, that is the result of being distinguished from dissimilar things' (TS 1008ab). Kamalaśīla provides the gloss, 'Because of the apprehension by persons who are mistaken by similarity with the entity which is excluded from things which are dissimilar' (TSP 1006–8b). This appears to be a case of what is called 'non-grasping of difference' (*bhedāgraha*). The idea here would seem to be that s_3 is said to be a cow because its mental representation, p_3, is more easily believed to be identical with its anyāpoha, $\sim -p_3$, than with any other mental occurrent apart from p_3 itself. We might say that one makes the identification $p_3 = \sim -p_3$ by disregarding the fact that the $\sim -$ function does not obey double negation.

It might seem somewhat unsettling to discover that one's knowledge that this is a cow rests on what could be called a case of mistaken identity. This would hardly seem odd to the Yogācāra-Sautrāntika, however. On his view, any subsumption of a particular under a general term must involve falsification, since the real particulars are absolutely unique. One can think of s_3 as resembling s_1 and s_2 in respect of cowness only through a process of mental construction which significantly alters what is given to us in our perception of the svalakṣaṇa. The trick, of course, is to give an account of

this process which allows for the possibility of successful practice in spite of the necessary element of falsification. The basic claim of the apohavādin is precisely that the ~ – function, when applied to mental images, yields a construct which both picks out just those entities which are called cows, and explains our failure to recognize that there is no feature which is common to those things called cows.

Does the apohavādin program succeed? Any critical estimate of the theory must confront the fact that this is a straight-forward case of nominalism by theft. That this is so was recognized by the originator of the theory, Diṅnāga, who boasted that the anyāpoha possesses all the properties of the universal, such as oneness and inherence in the many particulars (quoted by Kamalaśīla at TSP 1000). Presumably, one is to conclude that apohavāda is superior to the accounts of linguistic meaning provided by Nyāya and the Bhaṭṭa Mīmāṃsa by reason of 'lightness' or economy. This admission may, however, stir up certain misgivings. Has the apohavādin in fact avoided the illicit use of general entities in his explanation of linguistic practice? Dravid, it will be recalled, asserts that the negation of the opposite meaning is impossible unless the word already possesses a positive meaning. This complaint is a member of a family of objections which originates with Kumārila. What these objections all have in common is the suspicion that one or the other of the expressions 'not non-cow' and 'non-cow' must have as its meaning a real universal if the theory is to succeed.

Kumārila's basic objection is formulated in TS in the following fashion. If the determinacy of the meaning of 'cow' derives from the exclusion of those things which are not cows, then anyone who is able to use the word correctly must be in possession of some principle which may be used to distinguish between cows and non-cows. And there are only two ways in which such a differentiating principle might be formulated. On the one hand, it might be formulated in terms of similarities among those things which are cows; one could, on such a basis, determine that an entity is not a cow because it does not possess any of the features which are common to cows. On the other hand, such a principle might be formulated in terms of similarities among those things which are not cows; in this case one would determine that an entity is not a cow by noticing that it possesses features common to non-cows. In either case, however, it is assumed that there is something common to one or the other set of entities; and this 'something common' can be nothing other than a universal (TS 931–2).

The reply to this objection is interesting. What it comes to is an insistence that we distinguish between the intension and the extension of the expression 'non-cow'. The opponent has been misled by the occurrence of this expression in the statement, 'The meaning of "cow" is not non-cow'. He supposes that in order to employ the latter equivalence, we must know the meaning of 'non-cow'. In this case the theory must either fall victim to circularity or else stand convicted of assuming the existence of a universal. His supposition is false, however, for the 'non-cow' of this equivalence is an exclusion-class, not the meaning of a term. Put in terms of our limited domain, the meaning of 'cow' is $\sim -p_3$. Here the element corresponding to 'non-cow' is $-p_3$, and given the nature of p_3, this is determined as just the class, $\{p_4, p_5, p_6\}$. What would the meaning of 'non-cow' be? Let us assume that p_4 forms an exclusion-class different from that formed by p_5 and p_6, i.e., that

$$-p_4 = \{p_1, p_2, p_3, p_5, p_6\}$$
$$-p_5 = -p_6 = \{p_1, p_2, p_3, p_4\}$$

We might, for instance, suppose that s_4 is a goat, while s_5 and s_6 are sheep. In this way we are effectively barred from using any one of these three pratibhāsa individually in determining the meaning of 'non-cow'; and this is as it should be, for were we to do so we should be tacitly employing the assumption that there is something common to the non-cows. I would suggest that the meaning of 'non-cow' is $\sim -(-p_3)$ (or $\sim -(-p_1)$, etc.), that is, not non-(the exclusion class formed by a cow-pratibhāsa). To understand this expression correctly, we must first note that while prasajya-pratiṣedha apoha obeys double negation (TS 1153–7), it is nowhere said that paryudāsa apoha obeys double negation. I take this silence to mean that $- - p_3$ is not identical with p_3. In fact, if we consider the manner in which paryudāsa apoha operates, we can see that this is so. For $-(-p_3) = -\{p_4, p_5, p_6\}$, and this in turn is just $\{p_1, p_2, p_3\}$. Hence $\sim -(-p_3) = \sim \{p_1, p_2, p_3\}$, that is, what is *not* to be thought of as belonging to the class $\{p_1, p_2, p_3\}$. This yields the desired result, for $\sim \{p_1, p_2, p_3\}$ may truly be said to characterize the members of the class $\{p_4, p_5, p_6\}$, but is itself without existential import.

Several points emerge from this. First, in order to know the meaning of 'non-cow', we must know the meaning of 'cow'. That an entity is a non-cow is determined by the membership of its representation in the exclusion class $-p_3$. We can only be said to know that such an entity is properly called 'non-cow', however, if we know which pratibhāsa occur in the extension of

'cow'. Second, in order to know the meaning of 'cow' we must be able to determine the exclusion-class $-p_3$. We are capable of apprehending that s_3 is a cow only if we have some way of ascertaining which pratibhāsas are excluded by the occurrence of p_3. The third and most important point, however, is that this last task, the determination of the exclusion-class $-p_3$, does not require the employment of either of the two functions $\sim -p_3$ and $\sim -(-p_3)$. It is the nature of the pratibhāsa p_3 to exclude the class $\{p_4, p_5, p_6\}$. One does not need to know that the members of this class all belong to the extension of 'non-cow' in order to recognize that they are excluded by the occurrence of p_3. Nor must one be able to say that p_1 and p_2 are not excluded by the occurrence of p_3 in order for one to be aware that p_4, p_5, and p_6 are excluded. The occurrence of p_3 causes the exclusion of these pratibhāsa. Having learned the relevant linguistic convention, one then responds to the occurrence of this exclusion with the utterance, 'cow'.

One might, of course, wish to object to this reply on the grounds that it does not satisfactorily explain one's ability to apply the linguistic convention governing the use of 'cow' to this case. 'Surely', one wants to say, 'this convention must involve some rule for determining the nature of that exclusion-class which is to count as a criterion for the application of the word. And this rule must involve a property which determines the membership of the class. Thus to learn the convention is to learn the meaning of "non-cow".' In responding to this objection the apohavādin would want, I think, to make two distinct but related points. The first is that nowhere has it been assumed that there is anything common to the members of the class $\{p_4, p_5, p_6\}$. What one must learn to respond to in order to master the use of 'cow' is the occurrence of this class, not the occurrence of any feature or set of features which determines membership in this class. The second point is that one may learn to act in accordance with a rule without being able explicitly to formulate the rule in question. The apohavādin account of speech-production is a thoroughly causal account: A word is uttered given the occurrence of a pratibhāsa with certain causal properties, together with the disposition to respond to certain stimuli in some determinate fashion and the desire to engage in linguistic behavior (cf. TSP 1013–4). What this suggests is that the learning of the convention should be thought of as a kind of conditioning. When the situation is viewed in this light, the element of recognition is eliminated entirely from the mechanism which underlies the performance, 'This is a cow'. To learn the use of 'cow' is to acquire a

disposition to respond to the occurrence of the exclusion class $\{p_4, p_5, p_6\}$ with the utterance 'cow'. Surely one can do this without formulating the belief that the members of this class are the representations of just the non-cows; the psychological machinery is perfectly adequate to this task. Of course, once this machinery has done its work, we are in a position to say that s_4, etc., are non-cows; we are now able to recognize p_4, etc., as characterized by $\sim-(-p_3)$. We have this ability precisely because we are now operating at the linguistic or conceptual level. It hardly follows from this, however, that we must have employed this ability in ascending from the level of perception to the level of judgment. Still less does it follow from this that there must be something common to p_4, p_5, and p_6 by virtue of which they form an exclusion class.

This is, in any event, how I think the apohavādin wants to reply to this objection. Much more needs to be said on the subject, but so far there do not seem to be any glaring flaws in such a line of response. This is not of course the only objection which might be made to *apohavāda*, although it would appear to be the difficulty most frequently cited in the literature. Many other sorts of questions have also been raised, for instance concerning the ability of the theory to handle such logical terms as 'all', 'and', and 'not'. The defenders of the theory raise a number of interesting points in their discussions of these issues, but there is a great deal here which we do not yet understand. Also still unexplored is the question of the role which Diṅnāga's doctrine of categories plays in the theory. It would seem as if the categorial structure should serve to delimit the size of the exclusion classes, but it is not clear precisely how this would work. It is possible that further inquiry into these and other questions will ultimately unearth major defects in *apohavāda*. At this stage, however, and assuming that my reconstruction is at least in general outline correct, I can see no reason to believe that this brand of Buddhist nominalism is inconsistent. This result is of no little significance, for it is widely believed that a consistent extreme nominalism is unattainable. Here, as elsewhere, Dravid appears to have missed a feature of the Indian debate on the problem of universals which is of more than just historical interest.

I have left for last some remarks on the completeness of Dravid's survey of the problem of universals. There are a number of points on which our understanding would be enhanced by a more detailed examination of the problem. For instance, Dravid repeatedly refers to the apohavādins as 'the Buddhist nominalists', but it is clear that the Yogācāra-Sautrāntikas were not

the first Buddhist school to embrace nominalism. I suspect that much could be learned about *apohavāda* from an investigation of the attempts of the Vaibhāṣikas, early Sautrāntikas, and Theravādins to do away with abstract entities. That this problem has a long history in Buddhist thought can be seen from the fact that already at Kathā Vatthu VII.1 we find a debate over the question, whether dhammas (i.e., reals) can be collected under other dhammas. Likewise, there is much more to be said on the subject of Nyāya realism. Dravid reports the traditional Nyāya doctrine that the universal is ubiquitous (p. 93), but fails to mention the dispute among later members of the school over the meaning and the truth of this claim. He also notes the use which critics of realism made of the Nyāya distinction between universal (*jāti*) and imposed property (*upādhi*): If the great majority of abstract terms are the names not of real universals but of imposed properties, then why cannot such terms as 'cowness' be similarly construed? He sketches the replies of Uddyotakara and Kumārila to this objection (pp. 91–3), but there is, I believe, far more in the literature of the school on the question. Indeed, this is an area where further exploration might prove extremely rewarding, for the issue resembles in certain respects the contemporary debate over nominal terms and natural kind terms.

It would hardly be fair to criticize Dravid for failing to carry out such investigations, however, for he has managed to bring order to a rather enormous range of quite varied materials. Indeed, his work has the salutary effect of bringing to our attention just such gaps in our knowledge of the history of the problem. It also facilitates further exploration by providing an accurate picture of the general positions of the chief participants in the debate. In this respect Dravid has made a truly significant contribution to the field.

Dept. of Philosophy,
Illinois State University

NOTES

[1] Raja Ram Dravid, *The Problem of Universals in Indian Philosophy*. (Delhi: Motilal Banarsidass, 1972). Further references, by page number, will be incorporated in the text in parentheses.
[2] *Tattvasaṃgraha of Ācārya Śāntarakṣita With the Commentary 'Pañjika' of Śrī Kamalaśīla*, ed. Dwarkidas Shastri (Varanasi: Bauddha Bharati, 1968). References, by

verse number, to Tattvasaṃgraha (= TS) and to *Tattvasaṃgrahapañjika* (= TSP) will be incorporated in the text in parentheses.

[3] The terms 'verbally bound negation' and 'nominally bound neagtion' originate with Matilal. See Bimal K. Matilal, *The Navya-Nyāya Doctrine of Negation*. Harvard Oriental Series, Vol. 46, Cambridge: Harvard University Press, 1968, pp. 156–7.

[4] Cf. Matilal, *Negation*, pp. 92–3.

[5] This was pointed out to me by Prof. David Davis of the University of Hawaii. A similar point is made by Matilal in *Epistemology, Logic and Grammar in Indian Philosophical Analysis*, The Hague: Mouton and Co., 1971, p. 163.

This page intentionally left blank

Negative Facts and Knowledge of Negative Facts*

Brendan S. Gillon

INTRODUCTION

Negative facts have perplexed Western philosophers ever since the time of Plato.[1] But the philosophers of Europe and America have not been the only philosophers to have been perplexed by them; classical Indian philosophers too have pondered their nature. My interest here is to explore how the reflections of these classical Indian philosophers, transposed into the contemporary philosophical idiom, might enrich current metaphysical thinking about negative facts; and what I shall conclude is that at least one of these philosophers has a view of negative facts and knowledge of them, which, when so transposed, is very plausible indeed.

I shall begin by asking the fundamental ontological question of whether or not negative facts exist and then sketch various replies which European and American philosophers have given to it. Since these replies have not led to any decisive answer to the question, I shall then ask two other questions: the more specific ontological question of whether or not absences—surely paradigmatic examples of negative facts—exist; and the related epistemological question of what is known when the absence of something is said to be known. Answers to these questions comprise an important part of classical Indian philosophy; and I shall outline their answers to them, concluding that the most plausible answers to these questions are those of Jayanta Bhaṭṭa,

who maintained that absences do indeed exist and that they are known not only by inference but also by perception.

I NEGATIVE FACTS

Do negative facts exist or not? In one sense, of course, negative facts do exist. After all, if each of the following sentences should be true:

(1.1) There is not a hippopotamus in this room.
(1.2) Socrates is not alive.

(2.1) There is a dog in this room.
(2.2) Putnam is alive.

then, clearly, each of them expresses some fact.

This, however, is not what is meant by those who maintain that there are negative facts. What the proponents of the existence of negative facts mean is: first, that all facts can be partitioned (i.e. exhaustively divided) into two sets, positive and negative, neither of which, on its own, comprises the totality of facts making up the world; and, second, that facts of both kinds would obtain, even if there were no minds to know them.

Scepticism about the existence of negative facts is based, I believe, on two considerations. The first consideration is that 'negative facts are nowhere to be met with in experience'.[2] It is not clear, however, that this is so. Certainly common sense militates against this conclusion; for, common sense tells us that, if one simply looks into a refrigerator, and if there is broccoli in it, then one sees that there is broccoli in it, and if there is no broccoli in it, then one sees that there is no broccoli in it.

The sceptics' second consideration is that the division of facts into positive and negative ones is a spurious one, being an artefact of natural language. Looking again at the example sentences above, one might conclude that there is neither anything negative pertaining to the facts expressed in the pair of sentences in (1) nor anything positive pertaining to the facts expressed by the sentences in (2)—except, perhaps, that the first two facts are expressed by sentences with the adverb 'not' and the second two are expressed

by sentences without it. Some sentences, remarked upon by Frege,[3] make the point. As he observed, the very same thing can be expressed both by sentences with negative particles and by ones without them.

(3.1) Christ is immortal.
(3.2) Christ lives for ever.

(4.1) Christ is not immortal.
(4.2) Christ is mortal.
(4.3) Christ does not live for ever.

The conclusion, however, is not convincing, unless it can be shown, first, that every sentence expressed with a negative particle can be expressed by an equivalent one without any negative particle, and second, that the ontic commitments of the paraphrase are fewer than those of what is paraphrased. As I shall show below, it is difficult to meet these two desiderata jointly.

On the basis of the implicit assumption that every sentence with exactly one negative particle of the form N(S) has a version without the negative particle of the form S, Demos suggests that any sentence of the form N(S) can be paraphrased by a sentence of the form 'that which is inconsistent with S is true'.[4] For example, a sentence such as 'this rose is not yellow' can be paraphrased as 'that which is inconsistent with "this rose is yellow" is true' (or, more idiomatically, as 'that which is inconsistent with this rose's being yellow is true').

But, in the relation of inconsistency, what is inconsistent with what? Sentences as such are not inconsistent with one another, except insofar as they are expressions of things which are themselves inconsistent with one another. Thus, to answer the question just raised, one must first determine what sentences are expressions of. As Husserl makes clear in his *Logical Investigations*,[5] sentences are expressions, on the one hand, of propositions, which are the meaning correlates of sentences, and on the other hand, of states of affairs, which are the object correlates of sentences.[6] Inconsistency, then, is a relation either between propositions or between states of affairs. And so, to eschew acceptance of negative facts through a paraphrase of the kind suggested by Demos is to accept the existence of either propositions or states of affairs.

Russell, in his *Philosophy of Logical Atomism*,[7] rejects the existence of both. Propositions, he says, are not included in 'an inventory of the world', and so 'are not what you might call "real"'. And while states of affairs which obtain—that is, facts—are included in an inventory of the world, and hence are real; states of affairs which do not obtain are not in the inventory, and hence are not real. To these two explicit premises must be added this implicit one: real relations must have real relata—from which it follows that, if even one of the relata of a relation is not real, the relation itself cannot be real. Therefore, inconsistency cannot be a real relation between propositions, since propositions themselves are not real. Neither can inconsistency be a real relation between states of affairs, since at least one of a pair of states of affairs related by inconsistency cannot obtain, that is, cannot be a fact, and hence cannot be real. Thus, Russell concludes, 'it is simpler to accept negative facts as facts'.[8]

It might be thought that Russell's concerns could be met by supposing that inconsistency between a pair of states of affairs is supervenient upon inconsistency between actual properties ascribed to the very same actual individual substance or between actual relations ascribed to the very same actual individual substances taken in some fixed order. Thus, for example, the inconsistency between a state of affairs of a particular rose being red and a state of affairs of the very same rose being yellow, might be thought of as supervenient upon the inconsistency of the property of being red and the property of being yellow. Thus, to say that a particular rose is not yellow is to say that it has a property inconsistent with that of being yellow.

Alternatively, having discarded states of affairs as real and accepting only actual properties, relations, and substances as real, one might dispense with the Demos's paraphrase altogether and invoke one which mentions the relation of non-identity instead. Thus, to say that a particular rose is not yellow is to say that every property which this rose has is distinct from the property of being yellow.

These last two alternatives meet with one serious problem: *impersonalia*, that is, states of affairs expressed by uniclausal sentences with impersonal subjects (e.g. 'it is raining' and 'it is sunny'). As Reinach[9] has pointed out, such states of affairs can be inconsistent with one another, for example,

(5.1) It is sunny.
(5.2) It is overcast.

and their inconsistency cannot be recast as an inconsistency among constituents of factual states of affairs.

In short, either one accepts the existence of negative facts or one accepts inconsistency between either propositions or states of affairs. But, even if one accepts the latter alternative, one still must accept the existence of at least one negative fact, namely the fact that there is the negative relation of inconsistency.

It seems that the move to eliminate negative facts, or more generally, to eliminate negative states of affairs, founders on the need to invoke a relation which is itself negative. It might be thought that perhaps this predicament results from a misassignment of the burden of proof: the burden of proof lies, not with the opponents of negative facts, but with their proponents. In other words, it is not that one must show how negative facts can be eliminated, rather it is that one must show that there are reasons to posit their existence in the first place.

Russell, in his *Philosophy of Logical Atomism*, offers reasons to posit the existence of negative facts. Recall that Russell's logical atomism has, as a central tenet, that a logically ideal language, in displaying the logical form of the sentences of natural language, would thereby lead one to a correct ontology. According to Russell,[10] the atomic sentences of a logically ideal language which are true are rendered true by positive facts, while its false atomic sentences are rendered false by negative facts. In other words, the correspondence theory of truth needed for the logically ideal language requires two relations, correspondence, whereby facts render atomic sentences true, and what one might call counter-correspondence, whereby facts render atomic sentences false. Counter-correspondence is to be distinguished from a simple failure of correspondence. In the latter case, there simply is no fact to render a sentence true; in the former case, there is a fact to render the sentence false.[11]

Oaklander and Miracchi[12] have maintained that Russell did not have to posit the relation of counter-correspondence, and hence that he had no need to posit its factual relata of negative facts. They suggest that, instead, atomic sentences are rendered false,

not by counter-correspondence, but by a mere failure of correspondence.

The suggestion is a compelling one. Recall that the classical bivalent semantics for sentential logic is based on a truth-value assignment of exactly one of the truth-values of true and false to each of the atomic sentences of the language: that is to say, it is based on a total bivalent function from the set of atomic sentences to the truth-value set {T, F}. From this perspective, Oaklander and Miracchi's suggestion amounts to this: the total bivalent function from the set of atomic sentences to the truth-value set {T, F} can be replaced by a partial univalent function from the set of atomic sentences to the truth-value set {T}.

If the only task the relations of correspondence and counter-correspondence are required to do were to ground the truth-value assignment of atomic sentences, then the suggestion by Oaklander and Miracchi would be irreproachable. However, I do not believe that their only task is to ground the assignment of truth-values to atomic sentences. Surely it is necessary, in addition, to distinguish atomic sentences by what they express. Now, the differences among what true atomic sentences express can be grounded in the relation of correspondence: two true atomic sentences are expressively different by dint of their being made true by two different facts. But in what can be grounded the differences among what false atomic sentences express? Russell can ground the differences among what false atomic sentences express in counter-correspondence: two false atomic sentences are expressively different by dint of their being made false by two different (negative) facts. Oaklander and Miracchi's suggestion to give up the relation of counter-correspondence and the negative facts which go with it, leaves Russell with all false atomic sentences expressing the same thing, namely nothing.

Of course, the posit of negative facts and the relation of counter-correspondence is not the only way to address the problem of grounding the expressive difference among atomic sentences. An alternative is to posit a second relation of correspondence, call it expressive correspondence, whereby the expressive difference among atomic sentences is grounded in the differences among the states of affairs they express. On Russell's view, then, there are two verification relations, correspondence

and counter-correspondence, and two kinds of facts, positive (positive states of affairs which obtain) and negative (negative states of affairs which obtain). On the alternative view, there are two kinds of states of affairs, namely those which obtain and those which do not, and one verification relation, namely correspondence.

This alternative is unacceptable to Russell, since, as we saw before, Russell, the logical atomist, was unwilling to countenance the existence of states of affairs which are not facts. Yet Russell's fellow logical atomist, the early Wittgenstein, was perfectly willing to do so;[13] and, as a result, he can, and does, reject the existence of negative facts.[14] Thus, even with the assumptions peculiar to logical atomism, one is not compelled to accept the existence of negative facts; though, as Hochberg[15] points out, the acceptance of the existence of some counterpart for negative facts, such as the acceptance of the existence of states of affairs, is inevitable.

The reasoning just given for the acceptance of the existence of negative facts depends on assumptions peculiar to the doctrine of logical atomism; yet, even accepting these assumptions, one is still not compelled to embrace the existence of negative facts. The question arises: are there reasons which are not dependent on assumptions peculiar to a philosophical doctrine which nonetheless compel the acceptance of negative facts? I believe so; but they lie in an approach to the question of the existence of negative facts, different from the kind considered thus far.

II ABSENCES AND KNOWLEDGE OF ABSENCES

The approach is to consider paradigmatic instances of negative facts and see how such instances are to be accounted for within a larger metaphysical theory. A natural candidate for paradigmatic instances of negative facts is absences; and an obvious metaphysical setting within which to consider them is the epistemological setting of their knowability. Absences and their knowability has been well explored by classical Indian philosophers.[16] In what follows, I shall review the various metaphysical positions taken with respect to this pair of questions and defend the conclusion

of Jayanta Bhaṭṭa, namely that not only do absences exist but they can be known by mere perception.

Most classical Indian philosophers, like most contemporary philosophers, include perception and inference among the standard means of knowledge. Thus, for example, if I should enter a colleague's office and see him sitting at his desk, I know that he is sitting at his desk because of my visual perception of him there; and if, from my office window, I should see smoke rising from a building in the distance, I know that the building is on fire because of an inference from my observation of the smoke arising from it.

But now suppose that I enter a colleague's office and my colleague is not there. How do I know that he is not there? One obvious answer, tendered by the classical Indian philosopher Kumārila Bhaṭṭa (AD 620–680), is that I fail to perceive him in his office.[17] While this answer has the ring of common sense, there are reasons to baulk at accepting it. To begin with, perception and inference are, in one sense, faculties or cognitive capabilities, whereas non-apprehension (*an-upalabdhi*), the act of failing to perceive something, is not. There is, however, a more important reason to abandon the view that non-apprehension of something is a means of knowledge of its absence—something may be present yet not be perceived. Suppose that when I enter my colleague's office, it is night-time, the lights are out, and as a result his office is completely dark. Under these conditions, I would fail to see my colleague, even though he might very well be present.

There are essentially four possible metaphysical positions to account for what it is that I know when, after entering my colleague's office, I come to know that he is not present there. Each of the positions has been adopted and defended by certain classical Indian philosophers. On the one hand, some take the absence of my colleague from his office as a brute, negative fact. Of them, some hold knowledge of this fact to be perceptual; while others hold it to be inferential. On the other hand, some hold that the absence of my colleague from his office has no real ontic status at all, and believe that what there really is in the situation is just the sum of all the things present in the office. These latter philosophers hold that knowledge of my colleague's absence is just knowledge of what is present, though some believe the knowledge results from perception, while others from inference.

The first classical Indian philosopher known to have understood and addressed the problem which confronts the view of Kumārila Bhaṭṭa's is his older contemporary, Dharmakīrti (AD 600–660). According to Dharmakīrti, I know something, when I correctly report, having entered my colleague's office under the right conditions, that my colleague is absent from it; but what I know is not a negative fact and it is not known by perception. Rather, what I know is a simple (positive) fact and it is known by inference.

Recall that, for a philosopher like Demos,[18] the relevant inference is based on the relation of inconsistency between the simple (positive) fact I observe upon entering my colleague's office, on the one hand, and the state of affairs of my colleague's presence, on the other hand. To be sure, Dharmakīrti holds that the relation of inconsistency plays a role in many inferences which warrant one's claim to know some allegedly negative facts. For example, Dharmakīrti holds that from the presence of heat one can infer an absence of cold, since heat and cold are inconsistent (with respect to the same locus). But it is not his view that inference based on the relation of inconsistency plays a role in the case at hand. Rather, Dharmakīrti maintains, one reasons as follows: the causal conditions of perception known to me are such that, if my colleague were present in his office, I would see him; I do not see him; therefore, he is not present. The inference is counterfactual.[19]

One obvious objection to Dharmakīrti's position is one raised by Taylor[20] against Demos[21] and Russell.[22] Taylor argues that there is no more justification for the assertion that inference is involved in the case where one looks at a circle with no dot in it and observes that there is an absence of a dot from the circle, than there is for the assertion that inference is involved in the case where one looks at a circle with exactly one dot in it and observes that it has a dot. Taylor's point is that the two epistemic situations are phenomenologically the same in all relevant respects, and that since inference is not required to account for knowledge in the second case, it is not required in the first case either.

Taylor's point certainly seems to hold in the case at hand. Suppose I am standing at the threshold of my colleague's office door and looking into his office, my view unobstructed and the

room suffused with light. If my colleague is in the office and I report that he is, it seems that, if asked how I know, I can answer only that I see that he is there; and if my colleague is not in his office and I report that he is not there, it seems that, if asked how I know, I can answer only that I see that he is not there.

There are, however, other cases where the epistemic situation seems less phenomenologically similar than what is suggested by either Taylor's case or the case I just gave. Suppose that I drop a screw and suspect that it has fallen into a hole which I cannot see into but which I can put my hand into and completely search. Suppose that, after doing so, I report that the screw is not in the hole. If asked how I know, I would answer: if it had been there, I would have touched it; but since I searched for it by hand and did not touch it, it must not be there.

Indeed, given the right setting for the case where I report the absence of my colleague from his office, one can readily imagine situations in which inference does seem relevant to the determination of an absence, in the way it is relevant to the determination of the absence of the lost screw from the hole. Suppose that I and a companion have some reason to believe that my colleague might be in his office but hiding. If I were to go in and look about his office carefully and then report that he is not in the office, I might very well justify my report with counterfactual inference: I looked thoroughly; and if he had been there, I would have seen him.

It might be countered that the phenomenological dissimilarity in the epistemic situations of the knowledge of absences and knowledge of presences, evinced in these last two cases, results, not from the cognitive process at the time when the absence came to be known, but from the cognitive process at the time when pressed for a justification. If that is the case, then a deeper issue is now raised which undercuts Taylor's original argument: namely the issue of the accessibility of cognitive processes to introspection—an issue which space does not permit me to pursue here, as important as it is.

Taylor's objection aside, the question remains: does Dharmakīrti's counterfactual inference obviate the need to invoke the existence of absences? It seems not to. Suppose, for example, I know that my colleague is absent from his office. Let me symbolize this as $\neg p$. Now, according to Dharmakīrti, I know this, because

I have inferred it from two premises which I already knew: first, the counterfactual conditional that if my colleague were present in his office, then I would know that he is present there; and second, I do not know that he is present in his office. Let me symbolize them as $p > Kp$ and $\neg Kp$ respectively (where 'K' denotes the epistemic operator corresponding to the English 'a knows that' and '>' denotes the connective for counterfactual conditionals). But this gives rise to a dilemma: how do I know that I do not know that my colleague is present in his office? In other words, how do I know that $\neg Kp$? On the one hand, if Dharmakīrti holds that I perceive that I do not know that my colleague is present in his office, that is to say, that I perceive $\neg Kp$, then Dharmakīrti accepts thereby not only the existence of negative facts, in this case, negative mental facts; but also their perceptibility, both of which he wants to deny. On the other hand, if Dharmakīrti holds that I infer that I do not know that my colleague is present in his office, that is to say, that I infer $\neg Kp$, then there must be an inference to ground that claim, just as there is an inference to ground the initial claim that my colleague is not present in his office, that is, $\neg p$. An infinite regress of inferences becomes inescapable, despite Dharmakīrti's protestations to the contrary.[23]

A different position regarding negative facts and knowledge of them is the one adopted by the followers of Kumārila Bhaṭṭa. Though they maintained, unlike Dharmakīrti, that there are negative facts; nonetheless, they seem to have adopted his view that knowledge in such cases is derived only from counterfactual inferences.[24] Their view, of course, suffers from the very same dilemma just raised with respect to Dharmakīrti's view, namely the dilemma of having either to concede the perception of some negative facts or to accept, for any negative fact said to be known, an infinite regress of inferences of negative facts.

A younger contemporary of Kumārila Bhaṭṭa, Prabhākara (AD 650–720), who shared many of his philosophical assumptions, adopted the view that what I know, when upon entering my colleague's office I come to know that he is absent from it, is some visually perceived, positive fact, that is, some presence. But what would be the plausible candidate for such a presence? According to Prabhākara, it might be the presence of his chair. What else

could there be to the absence of my colleague from his chair than the presence of his chair? After all, as Buchdahl[25] has observed, one needs to draw only one picture to draw a picture either of my colleague's chair or of his chair without him in it. Yet, Prabhākara's metaphysical opponents were sceptical: they pointed out that, if my colleague's absence were identical with the presence of his chair, then anytime I perceive his chair, which includes the times that my colleague is in his chair, I should know that he is absent. This, according to the opponents, is patently absurd.

Prabhākara apparently sought to extricate himself from this objection by denying that the presence of my colleague in his chair is the same as the presence of his chair without him in it: in the former case, there are two presences, that of my colleague and that of his chair; but in the latter, there is only one, namely that of his chair. Opponents of Prabhākara suspected that this distinction cannot be made without directly or indirectly relying on an assumption of the existence of absences: for, as they pointed out, to distinguish the presence of my colleague in his chair from the presence of only his chair requires the use of such expressions as 'only' ('*eva*') or merely ('*mātra*'), which are implicitly negative expressions, since to say that there is only one presence is to say that there is at least one presence and there is no more than one presence.[26]

Prabhākara faced another objection. How is he to account for the fact that, when I see simply my colleague's chair, what I come to know is his absence and not one of the infinity of other absences which are also then present? One reply is to maintain that there is a difference between what I see and what I remark upon, and that this difference obtains equally in cases of the perception of what are unquestionably presences as well as in the cases of perception of what are allegedly absences. For the very same reason that when I look at my colleague's desk and remark that there is a pad of paper on it but fail to remark to myself that there is a pencil on it, I remark on his absence from his chair but I do not remark on the absence of the departmental chairman from it.[27]

An alternative to Prabhākara's position is one advocated by Jayanta Bhaṭṭa (fl. AD 890). He agrees with Prabhākara that perception is the relevant means for knowing that my colleague is absent from his office, but disagrees with him that all that is

relevant to my knowledge of his absence is my perception of some presence. Rather, Jayanta Bhaṭṭa holds that absences are brute facts unto themselves and that my colleague's absence in the case described above is known through perception. Jayanta Bhaṭṭa, like others of the same school, thought of perception as a causal process, the linchpin of which is contact between the object of perception and the appropriate sense organ. Thus, for example, in the tactile perception of the pencil in my hand, the key causal link is the contact of the organ, the skin, with the pencil. This view of perception, however, raised an immediate problem: if contact between an object of perception and the appropriate sense organ is the key link in the causal process of perception, how can there be perception of absences, for surely it is absurd to think of any sense organ as coming into contact with an absence.

Jayanta Bhaṭṭa's response to this objection is to revise the definition of sense perception: he restricted the above definition to the perception of positive facts and provided a special clause to account for perception of negative facts. According to Jayanta Bhatta, this move was not *ad hoc;* and at least within his overall metaphysical framework, it had independent motivation, for he believed that, though universals do not come into contact with the sense organs, they nonetheless can be perceived. In general, something is perceived just in case it is a physical thing suited to a sense organ and in contact with it, or it is a universal inhering in a physical thing and both the universal and the physical thing are suited to the sense organ, or it is an absence related to a place and both the absence and the place are suited to the sense organ.

To a contemporary philosopher, Jayanta Bhaṭṭa's response to this objection might seem both *ad hoc* and implausible. But let us not so readily discard his view of absences and their knowability. The implausibility of his response derives, not from his acceptance of negative facts and of their perceptibility, but from other assumptions, common to pre-modern thinkers, which modern metaphysics is coming to abandon. These assumptions are: that perception requires contact; that the relata of the causation relation are things; and that the mental relata of the relation of perceptual causation are the thing-like things, images. It seems to me that the most interesting aspects of Jayanta Bhaṭṭa's metaphysical views are retained, even when these assumptions are revised

into a form more congenial to current views of these metaphysical issues. Let us see how this can be done.

To begin with, in spite of its prevalence in the pre-modern world, the view that contact between the object of perception and the appropriate sense organ is the linchpin in the causal process resulting in perceptual knowledge, is simply untenable, except in some very attenuated sense of contact. Descartes is perhaps the last serious thinker to hold a contact theory of visual perception; and even in his theory, the contact between the eye and the object of vision is mediated by a long chain of things, each in contact with some other.[28] If one considers what is really essential in Jayanta Bhaṭṭa's epistemology, it is his view that perceptual knowledge results from a causal process; while contact between the object of perception and the appropriate sense organ plays an utterly ancillary role.

Another aspect of Jayanta Bhaṭṭa's metaphysics requiring some modification is his view of causation. To see what modification is required, we turn to the pioneering work by Zeno Vendler[29] on the expression of causation in natural language.[30]

We are all familiar with metonymy, the figure of speech whereby one word, typically a noun, literally denoting one thing, is used to denote something else, by dint of some well-known connection between the first thing and the second. Thus, for example, it is correct and idiomatic to say:

(6) The crown granted an amnesty to the rebels,

where the word 'crown', which literally denotes ornamental headgear worn by a monarch, is used to denote a monarch. Nor is metonymy a mere literary device, as the following sentences are intended to show.

(7.1) Have you read any (works by) Shakespeare?
(7.2) The White House (i.e. the President of the United States) has just made a statement.

Even entities which are not physical objects can be denoted metonymically. Thus, for example, what is forbidden by the first sentence below is not track shoes, but rather the activity of wearing track shoes in the locker room. Its literal paraphrase is given by the second sentence.

(8.1) Track shoes are forbidden in the locker room.
(8.2) The wearing of track shoes in the locker room is forbidden.

Metonymy also appears commonly and naturally in ordinary parlance pertaining to causation, as the following sentence and its literal paraphrase illustrate.

(9.1) This nail caused our flat tire.
(9.2) The puncturing of our tire by this nail caused it to go flat.

Metonymy, then, enables one to speak of a nail as a cause and a flat tire as an effect. Moreover, philosophical parlance easily abides such usage, as illustrated by the following remarks from Kant's *Critique of Pure Reason*:

> A room is warm, while the outer air is cool. I look around for the cause, and find a heated stove. Now the stove, as cause, is simultaneous with its effect, the heat of the room.[31]

In light of the role of metonymy in talk about causes and effects, it would be a mistake to permit such humdrum usage to inveigle one into believing that causation is a relation between physical objects. What, then, is the causation relation a relation between?

Since Vendler's insightful work, it is common to distinguish, at least *prima facie*, among physical objects, states of affairs, and events.[32] Events have both spatial and temporal location, since they take place or occur at a time and a place. Physical objects, however, have only spatial location, since they cannot be said to occur or take place. Facts have neither spatial nor temporal location.[33] What Vendler[34] shows is that, when metonymy is eliminated, causes and effects turn out to be states of affairs. And among the states of affairs which can serve as causes or effects are negative ones. The cause of an automobile accident, for example, might be a driver's not seeing a red light; and the driver's not seeing a red light, in addition, might be the effect of his being distracted by a passenger in his car. Moreover, absences are negative states of affairs, and they too can be both causes and effects. A negative state of affairs might be a cause: an absence of light in

a room might cause someone to stumble. A negative state of affairs might be an effect: a short-circuit caused there to be an absence of electrical power. And a cause and its effect might both be negative states of affairs: an absence of oxygen in a room might cause someone to go unconscious.

The final assumption to be revised pertains to the specific relata of perceptual causal processes. The widespread pre-modern model is that the relata are physical things, on the one hand, and images or mental pictures, on the other.[35] But this model simply won't do for propositional knowledge.[36] If a proposition is a proper characterization of what is the result of an act of propositional knowledge, then what is the object of that act? The usual, if implicit, answer is a state of affairs or a fact.[37] The current consensus, sometimes implicit, sometimes explicit, is that facts are the cause of an epistemic state, itself a fact or state of affairs, whose content is expressed by a proposition.

Once some of Jayanta Bhaṭṭa's metaphysical assumptions are reformulated to overcome their *prima facie* implausibility, a rather plausible account emerges of what it is that one knows when, in the right circumstances, one comes to know, say, of a colleague's absence from his office after having entered it, and how one knows what one knows in those circumstances: what one knows is the negative fact of a colleague's absence from his office and one knows it by means of perception. The plausibility of this account stands in contrast to the other three metaphysical alternatives, explored above, advocated by Prabhākara, Dharmakīrti, and Kumārila Bhaṭṭa. As was shown, the view of Prabhākara, namely that so-called negative facts are just presences of some kind, fails to say what, exactly, such presences are, without covertly relying on some form of negative fact; while the views of Dharmakīrti and the followers of Kumārila Bhaṭṭa, namely that only inference is a source of knowledge for whatever is known when an absence is said to be known, founder on the fact that the appropriate inferences must be grounded, inevitably, in a perception of something which is itself said to be an absence—which entails, contrary to their views, that at least some cases of knowledge of absences are cases of perception.

NOTES

* The first version of this paper was presented on 29 March 1990 to the American Philosophical Association, Pacific Division. A second version was presented on 26 March 1993 to the University of Ottawa's Department of Philosophy. I am grateful to the audiences on each occasion for their questions and comments. I would also like to thank Paul Forester, Richard Hayes, Graeme Hunter, Stephen Menn, Karl Potter and Mark Siderits for their suggestions and discussions of earlier versions of this paper.

1. *Sophist.*
2. Raphael Demos, 'A Discussion of a Certain Type of Negative Proposition', *Mind* 5: 26, 1917, p. 189.
3. Gottlob Frege, 'Negation', *Beiträge zur Philosophie des deutschen Idealismus*, 5: 1, 1919, pp. 143–57, English trans., Geach and Blackwell, 1952, p. 125.
4. Though Demos most frequently resorts to the word 'opposite' in his paraphrases, he makes it clear that both 'contrary' and 'inconsistent' are suitable equivalents. Demos, 'A Discussion of a Certain Type of Negative Proposition', p. 191.
5. Edmund Husserl, *Logische Untersuchungen*, Halle, Germany: M. Niemeyer, 1900, English trans., Findlay, 1970, Investigation I, ch. 1, esp. secs 11–13.
6. Bertrand Russell, 'The Philosophy of Logical Atomism', *Monist*, 5: 28–9, 1918, Reprint, ed. by Pears, 1985, Lecture III, p. 77.
7. Barry Smith, 'An Essay in Formal Ontology', *Gratzer Philosophische Studien* 5: 6, 1978, art. 2, discusses the connection of Frege's distinction between sense and reference, well known to Anglo-American philosophers, with this distinction of Husserl's.
8. Later, Bertrand Russell, *An Enquiry into Meaning and Truth*, New York: Allen and Unwin, 1940, pp. 73–4, 81–2, rejects, without argument, the existence of negative facts.
9. Adolf Reinach, 'Zur Theorie des negativen Urteils', in A. Pfänder, ed., *Münchener Philosophische Abhandlungen. Festschrift for Theodor Lipps*, 1911, sec. 12.
10. Russell, *The Philosophy of Logical Atomism*, Lecture III, p. 78.
11. See Russell, ibid., Lecture I, pp. 46–7; Lecture II, p. 51; and Lecture III, p. 71.
12. L. Nathan Oaklander and Silvano Miracchi, 'Russell, Negative Facts and Ontology', *Philosophy of Science*, 5: 47, September 1980, sec. III.
13. Ludwig Wittgenstein, *Logish-Philosophische Abhandlung: Annalen der Naturophilosophie*, 1921, trans., Pears and McGuiness (*Tractatus Logico-Philosophicus*), 1961, 2.04.
14. Ibid., 2.05 and 2.06.
15. Herbert Hochberg, 'Negation and Generality', *Noûs*, 5: 3, 1969, pp. 325–8.

16. A good overview of the various positions and arguments is Jayanta Bhaṭṭa's *Nyāya-Mañjarī*, end of the first *āhnika*. K.S. Varadacharya, ed., *Nyāya-Mañjarī of Jayantabhaṭṭa with Tippaṇi Nyāyasaurabha by the Editor*, 2 vols, Mysore: Oriental Research Institute, University of Mysore, 1969, v. 1, pp. 130–65; J.V. Bhattacharya, trans., *Nyāya-Mañjarī: The Compendium of Indian Speculative Logic*, New Delhi: Motilal Banarsidass, 1978, pp. 102–32.
17. See his *śloka-vārttika*, ch. 9 (Abhāva-vāda: On Negation), verse 11.
18. Demos, 'A Discussion of a Certain Type of Negative Proposition'.
19. See Dharmakīrti's *Nyāyabindu*, ch. 2, sūtras 12ff. (Dalshukh Malvania, ed., *Paṇḍita Durveka Miśra's Dharmottarapradīpa, Being a Sub-commentary on Dharmottara's Nyāyabinduṭīkā, a Commentary on Dharmakīrti's Nyāyabindu*, Patna: Kashiprasad Jayaswal Research Institute, 1955 [Tibetan Sanskrit Works Series: 2], 2nd edition, 1971, pp. 101ff); and his *Pramāṇa-vārttika*, ch. 3 (*Svārthānumāna*), verses 3–6 and his own commentary thereto Raniero Gnoli, ed., *The Pramānavārttikam of Dharmakīrti: The First Chapter with Autocommentary, Text and Critical Notes*, Rome: Istituto Italiano per il Medio ed Estremo Oriente, 1960 (Serie Orientale Roma: 23), pp. 4–6. This point is treated in detail by Brendan S. Gillon, 'Dharmakīrti and His Theory of Inference', in Matilal and Evans, eds, *Buddhist Logic and Epistemology*, 1986.
20. Richard Taylor, 'Negative Things', *The Journal of Philosophy*, 5: 49, 1952, pp. 443–5.
21. Demos, 'A Discussion of a Certain Type of Negative Proposition'.
22. Russell, *An Enquiry into Meaning and Truth*.
23. See his *Pramāṇavārttika*, ch. 3 (*Svārthānumāna*), v. 3 and the commentary thereto, Gnoli, *The Pramānavārttikam of Dharmakīrti*, pp. 4–5.
24. These philosophers, unlike Dharmakīrti, but like most other classical Indian philosophers, considered counterfactual reasoning (*tarka*) neither as a form of inference (*anumāna*) nor, in general, as a standard means of knowledge (*pramāṇa*). The case of counterfactual reasoning yielding knowledge of negative facts was set aside by them as a means of knowledge unto itself, called 'non-apprehension'. For further details, see Dhirendra Mohan Datta, *Six Ways of Knowing: A Critical Study of the Vedānta Theory of Knowledge*, Calcutta: Calcutta University Press, 1932, 2nd revised edition, 1960, Book III, ch. 4.
25. Gerd Buchdahl, 'The Problem of Negation', *Philosophy and Phenomenological Research*, 5: 22, 1961, pp. 176–7.
26. For linguistic confirmation of this view of 'only', see Laurence R. Horn, *A Natural History of Negation*, Chicago: The University of Chicago Press, 1989, p. 249.
27. This point has also been made by Buchdahl, 'The Problem of Negation', p. 175.
28. See, for example, his *La Dioptrique*.
29. Zeno Vendler, *Linguistics in Philosophy*, Ithaca, New York: Cornell University Press, 1967, chs 5 and 6; and Zeno Vendler, 'Causal Relations', *The Journal of Philosophy*, 5: 54, 1967.

30. Vendler's discussion is confined to English, but most of what he says holds with little or no modification of other languages.
31. Kant, *Critique of Pure Reason*, B 247–8, trans., Norman Kemp Smith.
32. It is customary in the specialist literature to use the term 'event' to include such entities as processes and states, due to a lack in English of an appropriate inclusive term.
33. See Vendler, *Linguistics in Philosophy*, ch. 5.12, for further discussion.
34. Ibid., ch. 6; and 'Causal Relations'.
35. Once again, see, for example, Hume's *An Enquiry Concerning Human Understanding*, secs IV, V and VII *passim*.
36. Russell (*The Philosophy of Logical Atomism*, Lecture IV, sec. 1, pp. 82ff), in his criticism of James and Dewey, exhibits an awareness of this problem.
37. See, for example, Alvin Goldman, 'A Causal Theory of Knowing', *The Journal of Philosophy*, 5: 64, no. 12, 22 June 1967, Reprint, ed. by Roth and Galis, 1970, p. 82.

REFERENCES

BHATTACHARYA, J.V., tr., *Nyāya-Mañjarī: The Compendium of Indian Speculative Logic* (Motilal Banarsidass, New Delhi, 1978).

BUCHDAHL, Gerd, 'The Problem of Negation', *Philosophy and Phenomenological Research*: v. 22, 1961, pp. 167–78.

CONRAD-MARTIUS, H., ed., *Gesammelte Schriften* (of Adolf Reinach) (M. Niemeyer, Halle, Germany, 1921).

DATTA, Dhirendra Mohan, *Six Ways of Knowing: A Critical Study of the Vedānta Theory of Knowledge* (Calcutta University Press, Calcutta, 1932), 2nd revised edition, 1960.

DEMOS, Raphael, 'A Discussion of a Certain Type of Negative Proposition', *Mind*: v. 26, 1917, pp. 188–96.

Dharmakīrti, *Pramāṇa-vārttika*, edn: Gnoli, ed., 1960. Partial English translation: Hayes and Gillon (tr.), 1991.

Dharmakīrti, *Nyāya-bindu*, edn: Malvania, ed., 1955. English translation: Stcherbatsky, 1930, vol. II.

FINDLAY, J.N., tr., *Logical Investigations*, 2 vols (Humanities Press, New York, 1970 [Internation Library of Philosophy and Scientific Method]). English translation of Husserl, 1990, 2nd revised edition.

FREGE, Gottlob, 'Negation', *Beiträge zur Philosophie des deutschen Idealismus*: v. 1, 1919, pp. 143–57. English translation: Geach and Black, trs, 1952, pp. 117–35.

GALE, Richard, *Negation and Non-being* (Basil Blackwell, Oxford, 1976

[American Philosophical Quarterly Monograph Series: Monograph No. 10]).

GEACH, Peter and Max BLACK, trs, *Translations from the Philosophical Writings of Gottlob Frege* (Basil Blackwell, Oxford, 1952), 2nd edition, 1960.

GILLON, Brendan S., 'Dharmakīrti and His Theory of Inference', in Matilal and Evans, eds, 1986, pp. 77–87.

GNOLI, Raniero, ed., *The Pramāṇavārttikam of Dharmakīrti: The First Chapter with the Autocommentary, Text and Critical Notes* (Istituto Italiano per il Medio ed Estremo Oriente, Rome, Italy, 1960 [Serie Orientale Roma: 23]).

GOLDMAN, Alvin, 'A Causal Theory of Knowing', *The Journal of Philosophy.* v. 64, no. 12 (22 June 1967), pp. 357–72. Reprint: Roth and Galis, eds, 1970, pp. 67–88.

HAYES, Richard P. and Brendan S. GILLON, 'Introduction to Dharmakīrti's Theory of Inference as Presented in *Pramāṇavārttika Svopajñavṛtti* 1–10', *Journal of Indian Philosophy:* v. 19, 1991, pp. 1–73.

HOCHBERG, Herbert, 'Negation and Generality', *Noûs:* v. 3, 1969, pp. 325–43.

HORN, Laurence R., *A Natural History of Negation* (The University of Chicago Press, Chicago, 1989).

HUSSERL, Edmund, *Logische Untersuchungen* (M. Niemeyer, Halle, 1900), 2nd revised edition 1913. English translation: J.N. Findlay, tr., 1970.

JAYANTA Bhaṭṭa, *Nyāya-Mañjarī* <End of Āhnika I>, edn: Varadacharya, ed., 1969. English translation: Bhattacharya, tr., 1978.

JHĀ, Gangānātha, tr., *Śloka-vārttika: Translated from the Original Sanskrit with Extracts from the Commentaries of Sucarita Miśra (The Kāśikā) and Pārthasārathi Miśra (Nyāyaratnākara), Bibliotheca Indica,* 1900, nos 965, 986, 1017, 1055, 1091, 1157 and 1183.

——, tr., *Prabhākara School of Mīmāṁsā* (Allahabad University, Allahabad, 1911), Allahabad University Series: v. 1.

KUMĀRILA Bhaṭṭa, *Śloka-vārttika,* edn: Śāstri, ed., 1978. English translation: Jhā, tr., 1900.

MALVANIA, Dalshukh, ed., *Paṇḍita Durveka Miśra's Dharmottarapradīpa, Being a Sub-commentary on Dharmottara's Nyāyabinduṭīkā, a commentary on Dharmakīrti's Nyāyabindu* (Kashiprasad Jayaswal Research Institute, Patna, 1955), Tibetan Sanskrit Works Series: 2, 2nd edition, 1971.

MATILAL, Bimal K. and R.D. EVANS, eds, *Buddhist Logic and Epistemology* (D. Reidel Publishing Co., Dordrecht, 1986).

OAKLANDER, L. Nathan and Silvano MIRACCHI, 'Russell, Negative Facts, and Ontology', *Philosophy of Science:* v. 47, no. 3 (September 1980), pp. 434–55.

PEARS, David, ed., *The Philosophy of Logical Atomism* (Open Court, La Salle, 1985 [Open Court Classics]).

PEARS, David and B.F. MCGUINNESS, trs, *Tractatus Logico-Philosophicus* (Routledge and Kegan Paul, London, 1961), 2nd impression, 1963.

PFÄNDER, A., ed., *Münchener Philosophische Abhandlungen. Festschrift für Theodor Lipps* (Barth, Leipzig, 1911).

PRABHĀKARA, *Bṛhatī*, edn: Śāstri, 1934. English translation: Jhā, tr., 1911.

REINACH, Adolf, 'Zur Theorie des negativen Urteils', in A. Pfänder, ed., 1911, pp. 196–254. Reprinted in H. Conrad-Martius, ed., 1921, pp. 56–102. English translation: Smith, Barry, tr., 1982.

ROTH, Michael D. and Leon GALIS, eds, *Knowing: Essays in the Analysis of Knowledge* (University Press of America, Lanham, 1970), 2nd edition, 1984.

RUSSELL, Bertrand, 'The Philosophy of Logical Atomism', *Monist:* v. 28, 1918, pp. 495–527; v. 29, pp. 32–63, 190–222, 345–80. Reprint: Pears, ed., 1985.

——, *An Enquiry into Meaning and Truth* (Allen and Unwin, New York, 1940).

ŚASTRI, Svāmī Dvārikādāsa, ed., *Ślokavārttika of Śrī Kumārila Bhaṭṭa: With the Commentary Nyāyaratnākara of Śrī Pārthasārathi Miśra* (Tara Publications, Varanasi, 1978), Prāchyabhārati Series: v. 10.

ŚASTRI, S.K. Rāmanātha, ed., *Bṛhatī: edited with Śālikanātha Miśra's* Ṛjuvimālapañcikā and Bhāṣyapariśiṣṭa, 3 parts (University of Madras, Madras, 1934), Madras University Sanskrit Series.

STCHERBATSKY, Th., *Buddhist Logic*, 2 vols (The Academy of Sciences of the USSR, Leningrad, 1930; rpt, Dover Publications, New York, 1962 [Bibliotheca Buddhica: 26]).

SMITH, Barry, 'An Essay in Formal Ontology', *Gratzer Philosophische Studien:* v. 6, 1978, pp. 39–62.

——, tr., 'On the Theory of Negative Judgement', in Barry Smith, ed., 1982, pp. 315–77. English translation: Adolf Reinach, tr., 1911.

——, ed., *Parts and Moments, Studies in Logic and Formal Ontology* (Philosophia Verlag [Analytica], Munich, 1982).

TAYLOR, Richard, 'Negative Things', *The Journal of Philosophy:* v. 49, 1952, pp. 433–49.

VARADACHARYA, K.S., ed., *Nyāya-Mañjarī of Jayantabhaṭṭa with Tippaṇi Nyāyasaurabha by the Editor*, 2 vols (Oriental Research Institute, University of Mysore, Mysore, 1969).

VENDLER, Zeno, *Linguistics in Philosophy* (Cornell University Press, Ithaca, New York, 1967a).
——, 'Causal Relations', *The Journal of Philosophy.* v. 54, 1967b, pp. 704–13.
WITTGENSTEIN, Ludwig, *Logish-Philosophische Abhandlung. Annalen der Naturphilosophie,* 1921. Translation: Pears and McGuinness, trs, 1961.

MEREOLOGICAL CONSIDERATIONS IN VASUBANDHU'S "PROOF OF IDEALISM"*
(Vijñaptimātratāsiddhiḥ)

Matthew Kapstein

And the venerable Nāgasena said to Milinda the king: "You, Sire, have been brought up in great luxury, as beseems your noble birth. If you were to walk this dry weather on the hot and sandy ground, trampling under foot the gritty, gravelly grains of the hard sand, your feet would hurt you. And as your body would be in pain, your mind would be disturbed, and you would experience a sense of bodily suffering. How then did you come, on foot, or in a chariot?"

"I did not come, Sir, on foot. I came in a carriage."

"Then if you came, Sire, in a carriage, explain to me what that is. Is it the pole that is the chariot?"

"I did not say that."

"Is it the axle that is the chariot?"

"Certainly not."

"Is it the wheels, or the framework, or the ropes, or the yoke, or the spokes of the wheels, or the goad, that are the chariot?"

And to all these he still answered no.

"Then is it all these parts of it that are the chariot?"

"No, Sir."

"But is there anything outside them that is the chariot?"

And still he answered no.

"Then thus, ask as I may, I can discover no chariot. Chariot is a mere empty sound. What then is the chariot you say you came in? It is a falsehood that your majesty has spoken, an untruth! There is no such thing as a chariot! You are king over all India, a mighty monarch. Of whom then are you afraid that you speak untruth?" And he called upon the Yonakas [Greeks] and the brethren to witness, saying: "Milinda the king here has said that he came by carriage. But when asked in that case to explain what the carriage was, he is unable to establish what he averred. Is it, forsooth, possible to approve him in that?"

---

* I wish to thank Professors A. L. Basham, Roderick M. Chisholm, and James Van Cleve for their advice and criticism in connection with this paper.

When he had thus spoken the five hundred Yonakas shouted their applause, and said to the king: "Now let your Majesty get out of that if you can?"

And Milinda the king replied to Nāgasena, and said: "I have spoken no untruth, reverend Sir. It is on account of its having all these things—the pole, and the axle, the wheels, and the framework, the ropes, the yoke, the spokes, and the goad—that it comes under the generally understood term, the designation in common use of 'chariot.'"[1]

Some two millenia ago, in northwestern India, the Buddhist sage Nāgasena and the Greek king Menander came together to discuss problems of philosophy and religion. In the passage reproduced above, their dialogue enters a territory which has continually given rise to philosophical puzzlement within both the Indian and Hellenic traditions and their offshoots: How are we to understand the mereological structure of matter? What is the relation of the whole to its parts? The answer which Menander finally produces in response to Nāgasena's Socratic questioning is recognizably an early version of a theory that has not been without its modern proponents: wholes are not entities that exist in addition to their parts; they are merely logical constructions.

In the present paper I shall not be concerned with this ancient logical constructivism *per se*, though I shall seek to indicate something of the manner in which it formed the conceptual background for the development of Vasubandhu's "proof of idealism," which belongs to a more developed phase of Buddhist mereological thought. Before, however, turning to the arguments themselves, some introductory remarks of a general nature are called for.

Vasubandhu, the founder of Buddhist philosophical idealism, has been the object of considerable philological research since the latter part of the nineteenth century. Historical data alone justify the efforts that have been made, for, with the translation of his works into Chinese, Tibetan, and not a few other languages, Vasubandhu's thought became, from about the sixth century onward, a dominant force in central and east Asian intellectual life. Moreover, his substantial contributions to logic, epistemology, and metaphysics are more than sufficient to justify his prominent position in the history of Indian philosophy, even if his impact abroad were totally disregarded. Still, for the philosopher, the relative lack of attention paid by philologists to Vasubandhu's actual philosophy remains an obstacle to study. What Jonathan Barnes has said of the Presocratics may be restated, *mutatis mutandis*, in the present instance:

> . . . if the linguistic expression and the historical context of the Presocratics have been exhaustively discussed, the rational content of their thought has been less thoroughly scruntized. By and large, scholars have have asked what the Presocratics said, and what external circumstances

may have prompted their sayings; they have not asked whether the Presocratics spoke truly, or whether their sayings rested on sound arguments.[2]

Two particular difficulties stand in the way of our attempt to formulate a philosophical reconstruction of Vasubandhu's teaching, even after philology makes clear what, exactly, he said.[3] First, he was, like many another classical Indian writer, an elliptical stylist to an extreme degree: Vasubandhu assumed himself to be writing for the beneficiaries of a thorough Buddhist scholastic education in Sanskrit, and so presupposed an audience for whom much could be left unstated. Today, we must fill in the gaps to the best of our ability by relying upon the entire corpus of his surviving work, the writings of others in his tradition, and the demands of reason. In the second place, Vasubandhu, perhaps in part because he was engaged in the difficult task of forging new philosophical tools, does not always achieve the perspicuity for which he certainly aims and which marks the work of several of his philosophical successors.[4] Often, for instance, he will voice an intuition in terms of a concrete example; and it thus remains for the reader to penetrate to the abstract reasoning beneath the surface. With this in mind, then, I offer here an exposition and interpretation of the rational content of Vasubandhu's "proof of idealism."[5]

Let us begin by returning to the conversation between Nāgasena and Menander. The following mereological principles are certainly implied there:

(P1) For all x and all y, if x is a composite material whole and y is a proper part of x, then it is not the case that $x = y$. [Is it the pole that is the chariot?]

(P2) For all x and all $y_1, y_2, \ldots$, if x is a composite material whole and $y_1, y_2, \ldots$ are the proper parts of x, then it is not the case that x = the *mere sum* of $y_1, y_2, \ldots$. [Then is it all these parts of it that are the chariot?]

(P3) For all x and all $y_1, y_2, \ldots$, if x is a composite material whole and $y_1, y_2, \ldots$ are the proper parts of x, then it is not the case that there exists some z such that (i) $x = z$ and (ii) z is discrete from $y_1, y_2, \ldots$. [But is there anything outside them that is the chariot?]

One possible objection to this scheme may be immediately raised when we consider (P2) alone. True enough, one might say, the whole is no *mere sum* of its proper parts. But is it not, rather, a unified, ordered collection of those

parts? One who advances this objection might wish to affirm the following to supplement (P2):

> (P4) For all x and all $y_1, y_2, \ldots$, if x is a composite material whole and $y_1, y_2, \ldots$ are the proper parts of x, then x = the unified, ordered collection constituted by all and only $y_1, y_2, \ldots$. [It is on account of its having all these things . . . that it comes under the generally understood term . . . "chariot."]

Now, if a *mere sum* and a *unified collection* are collections of just the same parts a, b, and c, then what is it that distinguishes them? One way to answer this question is to assume that there is some relation, say of *strong conjunction*, which is to be differentiated from the weaker notion of logical conjunction. With reference to the latter it is true that for all x and all y, where x and y are discrete, concrete, individual things, there is a *mere sum* which is the *logical conjunction* of x and y. Assuming now the (undefined) relation of *strong conjunction*, we can say that a, b, and c, in the example above, form a *mere sum* iff *strong conjunction* does not obtain among them; and that they form a *unified collection* iff it does so obtain.

How are we to respond, though, to one who points out that in what has been said so far there is nothing to preclude each of the parts of some whole from being itself a unified collection of parts, each of which is itself a unified collection, and so on, *ad infinitum*? In that case, any material whole must actually have infinitely many material parts, which seems absurd. One hypothesis which avoids this difficulty by denying that any unified collection actually has infinitely many material parts is atomism: namely, the theory which holds that any unified collection actually has only a finite number of material parts. These parts must be thought to be spatially extended (being material), simple (for they do not themselves have parts), and empirically indivisible (for, if empirically divisible, then they must have parts). Let us call such atoms *minimal parts*. And let us further stipulate that for all x, where x is a composite material whole, there is some set S of minimal parts such that (i) S has a finite number of members and (ii) each member of S is discrete from every other. Then we may restate our four mereological principles as *principles of minimal part atomism* (Pa), thus:

> (Pa1) For all x and all S, if x is a composite material whole and S is the set of the minimal parts of x, then it is not the case that x = any member of S.

(Pa2) For all x and all S, if x is a composite material whole and S is the set of the minimal parts of x, then it is not the case that x = the *mere sum* of all members of S.

(Pa3) For all x and all S, if x is a composite material whole and S is the set of the minimal parts of x, then it is not the case that there exists some z such that (i) $x = z$ and (ii) z is discrete from all members of S.

(Pa4) For all x and all S, if x is a composite material whole and S is the set of the minimal parts of x, then x = the unified, ordered collection constituted by all and only the members of S.

These four principles have been implied, if I am not mistaken, by most of the ancient schools of atomism. They admirably preserve many of our ordinary intuitions about the material world; and they burden our ontology with only one putative entity, the minimal part, and one putative relationship, the relation of strong conjunction. Minimal part atomism is thus attractive to some because it retains the metaphysical simplicity of logical constructivism, while refusing to countenance the absurd notion that corporeal substances might be actually composed of an infinite number of extended parts. Still, not a few ancient philosophers found atomism to be a flawed doctrine. Zeno, Aristotle, and Vasubandhu agree in rejecting various theories which explicitly or implicitly embrace minimal part atomism, but their alternatives are, respectively, monism, the theory of extended continua, and idealism.

It is instructive to ponder the considerations which led Aristotle to shun idealism in favor of his continuity theory and Vasubandhu to make just the opposite move. For the Stagirite, the notion that a body might be "nothing but an appearance" is patently absurd and is tantamount to maintaining that "its constituents are nothings . . . it might both come-to-be out of nothings and exist as a composite of nothings."[6] If we agree that this is not possibly correct, and reject atomism at the same time, then we have no choice but to seek, with Aristotle, an alternative theory of matter.

Vasubandhu, on the other hand, aims to demonstrate that the atomic theory is *both* false *and* necessary: "atomic distinctions must be supposed; and there can be no simple atom." Wherefore, matter is naught but ideal. Vasubandhu, if he is to make his case, must demonstrate the truth of the premises of a simple *modus tollens* argument:

(1) If material things exist independently of the perceptions in which they are given, then they must be atomic in composition.

(2) But they cannot be atomic in composition.

Therefore, (3) material things do not exist independently of the perceptions in which they are given.[7]

This becomes a "proof of idealism" if supplemented by the following definition:

x is ideal = Df x does not exist independently of the perception(s) in which it is given.

The main body of Vasubandhu's argument (sections ii and iii) is devoted to the demonstration of premises (1) and (2). This, however, is preceded by a summary statement of the alternatives to idealism and their defects (section i). The alternatives enumerated are:

(*i*) the whole is a simple "part-possessing form" (*avayavirūpa*);

(*ii*) it is a *mere sum* of atomic parts; and

(*iii*) it is a *unified collection* of atomic parts.

It is important to consider how it is that Vasubandhu intends these to be understood, and how they are to be related to the main body of the argument which follows.

(*i*) The "part-possessing form."

Against the early Buddhist logical constructivism represented in such texts as the *Questions of King Milinda*, the Hindu philosophers of the Nyāya-Vaiśeṣika school posited that the whole is something in addition to the unified collection of its parts.[8] There are two lines along which their doctrine might be interpreted: both the unified collection of parts and the whole actually exist; or only the whole actually exists. The latter line of interpretation suggests an analogue to the Aristotelian continuity theory and is taken up by Vasubandhu in section iii., where he challenges the doctrine of simple, extended wholes. In section i., however, his target is the notion that the whole exists above and beyond the actually existing parts. Vasubandhu maintains that such a whole is "not apprehended," i.e., its concept is in some sense counterintuitive. How so? Certainly, when we see, e.g., a car-

riage, we do not see those parts of the carriage which are visible to us and the carriage besides. But I suspect that Vasubandhu intends something more than that, and what it might be I shall try to make clear in what follows.

Assume a given whole to be the conjunction of a unified collection of parts and something x existing above and beyond that collection. This may commit us, of course, to abandon (P3) and its atomist analogue (Pa3). If x is an *abstractum*, then it is difficult to see in what sense it actually *exists* in addition to the unified collection of parts. Indeed, the doctrine of the part-possessing form would, in that case, be nothing but a fancy sort of logical constructivism. If, on the other hand, x is a *concretum*, is it simple or complex? If simple and existing in addition to the unified collection, then we are indeed committed to abandoning (P3)/(Pa3). Now, empirical considerations aside, x obviously can, under these conditions, be thought of as a simple entity which is logically independent of the unified collection. As such, it merits no special discussion apart from the doctrine of simple, extended wholes in section iii. But, if it were deemed *complex* and existing in addition to the unified collection, then the entire mereological puzzle is merely reintroduced with reference to x, and nothing whatever has been gained by its supposition. Finally, should it be assumed that there is a part-possessing form, but that it does not exist in addition to the unified collection, then Vasubandhu is quite correct in asserting that it is nothing discrete from the minimal parts of the unified collection. In short, Vasubandhu seems to maintain that if the hypothesis of the part-possessing form has any substance to it at all, it can only be according to the continuity theorist's reading of it, and that will be taken up below.[9]

(*ii*) The whole as mere sum.

Vasubandhu's rejoinder to the proponents of this second view—"single atoms are not apprehended"—suggests that he considered this alternative, too, to be counterintuitive. The thoughts which led to our distinguishing mere sums from unified collections and positing the relation of strong conjunction do lend support to Vasubandhu's intuition. Still, the assertion that a doctrine which holds that wholes are mere collections of minimal parts entails our being able to perceive minimal parts *directly* may seem to be excessive. But consider: if wholes *are* mere sums and not even unified collections, then in fact there *is* no difference between what we ordinarily take to be mere sums and wholes. Nonetheless, there *appears* to be some difference between them. So, wholes are only apparent things, that is, they do not exist independently of the perceptions in which they are given. Thus, even if we do posit

noumenal atoms of some sort, as long as they do not form unified collections, we must accept some form of idealism, at least with respect to apparent wholes, which is just what Vasubandhu set out to prove in the first place.

(*iii*) Wholes as unified collections of atoms.

This thesis is to be given thorough consideration in section ii. Note that at this point in the argument, we are left with two viable theories of matter: atomism and the continuity theory. The refutation of these will give Vasubandhu the premises he seems to require.

Section ii. Contra atomism.

The refutation of atomism is in two parts: ii.a., in which Vasubandhu argues that the supposition of the conjunction of partless minimal parts entails various absurdities; and ii.b., where he seeks to demonstrate that spatial extension and simplicity are, in any case, mutually exclusive properties. It appears that both arguments are intended to support one and the same conclusion, namely, that the atomist's concept of simple minimal parts is defective and so cannot be made to account for complex wholes.

Section ii.a. requires, I think, that we suppose Vasubandhu to have intended some such argument as:

(1) If there exists an atom which is a simple substance, it must be able to enter into conjunction with other atoms.

(2) But it is not possible for it to do so.

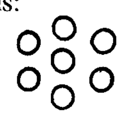
Therefore, (3) there does not exist an atom which is a simple substance.[10]

The first premise, which requires the atomic theory to posit atoms that do form unified collections, and not only mere collections, has been already established. Only the second premise need now be proved. To do so, Vasubandhu presents the atomist with two dilemmas: one arising from the notion of minimal parts conjoining directly with one another, the second arising from an alternative "molecular" theory of conjunction.

ii.a1. The first dilemma of conjunction.

Consider one partless atom entering into conjunction with six others. Either

all occupy discrete spaces, thus:



or else they all occupy the very same space:



"To occupy" a space means here "to fill" that space, and not simply "to fall within" it.

If we grasp the first horn of the dilemma, then surely we are committed to the view that our atoms have parts, for they have discrete parts in contact with discrete atoms. Should we opt for the second alternative, however, we must admit that all the atoms in the group must collapse into the same atom-sized space.

The above, or something close to it, is what I take Vasubandhu to have in mind here. I confess that while the problem posed by the first horn of the dilemma seems to me to be clearly a real difficulty, I am by no means confident in the interpretation of the second. Let us examine it somewhat more carefully.

Is it possible that Vasubandhu had in mind here not extended minimal parts, but unextended points? In that case, the dilemma created in ii.a1. would have us choose between extended atoms, which must have parts in order to conjoin, and unextended material points, an infinite number of which might "conjoin" without any space being thereby filled. However, I doubt very much that Vasubandhu is entertaining here the idea of there being unextended material points, an omission which, though partially remedied below, might prove to be devastating for his "proof of idealism."[11] At the same time, should we adhere to the first interpretation proposed, which in any case better accords with Vasubandhu's own words, it may be well to ask whether or not the notion of several atoms occupying the very same space is, in fact, problematic. Do the "molecules" formed of the compresent atoms conjoin, or not? What is to prevent all the atoms in the universe from occupying the same space? Is it possible for any two, or more, material things to occupy one and the same space? The second horn of the dilemma calls upon us to consider these questions.

Vasubandhu's own answer to the last question would have certainly been "no."[12] It is one of the properties of matter, as we ordinarily understand it, that no two bits of it can occupy simultaneously the very same space. Any

theory that denies this fundamental intuition must be prepared to explain just why it is, then, that *all* the matter in the universe does not simply collapse into the same space. Moreover, should such a theory in some way limit the quantity or number of extended substances that may be compresent, it will be faced with the first question posed above: for, if the "molecules" do not conjoin with one another, we confront again the problem of mere collections; and if they do conjoin, we are left with the puzzle of conjunction itself, but this time with reference to atom-sized clusters, rather than atoms. As Vasubandhu's second dilemma presents that puzzle all over again, in fact with reference to "molecules" and not atoms, it seems not unlikely that he well realized that the second horn of the first dilemma might be taken to be the lesser of the two evils offered, the point of departure for a molecular theory requiring clusters of atoms to serve as the finest actual parts of coarse matter; though, as seen above, he would not have countenanced the idea that the atoms making up those clusters might be literally compresent.

It was the Vaibhāṣika school of Buddhist realism that, with remarkable genius, developed a molecular theory not clearly susceptible to objections such as those raised in the first dilemma of conjunction; and it is this theory that Vasubandhu attacks in section ii.a2.[13] The Vaibhāṣika theory in its developed form seems to have required unextended points of resistance, indivisible even in thought, falling within, but not filling, given spaces, and at the same time capable of existing only in clusters. These clusters conjoin to form gross matter. This, according to Vasubandhu, merely introduces a new dilemma for the atomist: the cluster must be either the same as, or different from, the atoms which compose it. If it is different, then of what is it a cluster? On the other hand, if it is the same, then for what reason should we suppose it to be capable of conjunction at all, when its atomic parts are not? Surely Vasubandhu is right to object to the notion that the atomic clusters might be discrete from the atoms of which they are composed. Aggregations that are not aggregations of anything are, like dollars worth no cents, meaningless fictions. Assuming, then, that the aggregations are mere collections of unconjoined atoms, can we assume unified collections ever to be generated from them? Vasubandhu found such an assumption wholly untenable. Some contemporary philosophers would, I think, not be so ill-disposed to it.[14] Their views, however, would have to be examined in the light of modern physical theory.

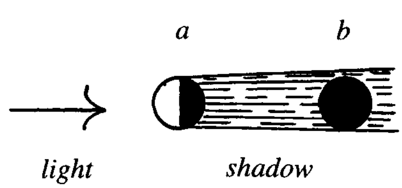
It should be noted, though, that a cogent point-particle theory would mitigate severely the force of Vasubandhu's argument: as long as such a theory cannot be decisively rejected, Vasubandhu's "proof of idealism" is, in fact, no proof at all. Nonetheless, it should not be forgotten that Vas-

ubandhu aimed to challenge premodern minimal part atomism, not twentieth century particle physics. His conclusion to section ii.a., that his pair of dilemmas has shown the atom to be incapable of conjunction and that, hence, the existence of the minimal part cannot be established, is not at all unwarranted, given the ancient theories upon which he directed his assault.

Section ii.b. Extension and simplicity.

The second argument against atomism is understood to be entirely independent of the first. It is also the argument one most frequently encounters in Western criticisms of atomism: spatial extension and simplicity cannot be properties of one and the same substance.[15] As Vasubandhu plainly puts it, the atom, if spatially extended, "has a front part that is different from its bottom part."[16] It may be recalled that, in formulating the concept of the minimal part, all that was required in addition to extension and simplicity was *empirical indivisibility*. Vasubandhu and most other opponents of minimal part atomism object that as long as the logical possibility of division remains, that is to say, as long as we can conceive of the atom as having parts, the minimal part cannot be simple. This claim will have to be reexamined below in connection with the question of extended, simple wholes, for the notion that having possible parts entails having actual parts is just what the continuity theorist will deny. Still, the objection does succeed in pointing out a certain arbitrariness in the atomist's bringing the process of division to a halt with hypothetical atoms. We are driven, in effect, to choose between the apparent absurdity of infinite division and the continuity theory.

At this point in the text, Vasubandhu adds a peculiar argument concerning shade and shadow. His intuition may be represented with the aid of a diagram:



Consider two particles, *a* and *b*, such that *a* is struck by light and *b* is in the shadow of *a*. If we think of *a* as an extended thing, then we must think of it as having a part that is illumined and a part that is shaded. But if we

assume it to be partless this cannot be: the light must either stop short of *a*, or else it must penetrate *a* through and through. If the former, we must explain the light's strange behavior; if the latter, then there is no reason to suppose that the light does not reach *b* as well.

It is obvious that this example is intended only to illustrate the intuition that extended things have parts and, as such, it remains susceptible to the continuity theorist's objections. Moreover, the point-particle theorist can maintain without absurdity that *a* might be an unextended point of resistance at which the light simply terminates, and so need not be troubled by Vasubandhu's example. That Vasubandhu did not seriously consider the implications of the point-particle hypothesis is almost certainly indicated by his equation, in this passage, of resistance with extension: "if no atom whatever has discrete parts, such as might obstruct passage from one place to another, then, in the absence of such obstruction, owing to the co-spatiality of all, an entire cluster must be a mere atom." If atoms are extended, they must have parts; if unextended, the universe is swallowed in a mere point.[17]

Section ii.b. closes with a response to the philosopher who supposes the phenomena of shade and shadow to depend upon atomic clusters, rather than upon single atoms. That supposition requires the clusters to be discrete from the atoms which compose them; otherwise there is no reason to suppose these phenomena do not depend on the atoms themselves.

Section iii. Against continuity.

The abandonment of the atomic theory leaves intact the data supplied by our physical senses: "So long as the defining characteristics of material wholes are present, does it matter whether their concepts correspond to atomic collections, or not?" After he has denied the atomic theory, the postulate that wholes are complex substances composed of minimal parts, another possibility remains; namely, that wholes are simple, extended continua, possibly divisible, but not composed of *actual* parts. This, of course, is the notion at the heart of both Aristotelian and Nyāya-Vaiśeṣika mereological speculation, a way to avoid the terrible conclusion that a body might be "nothing but an appearance." Vasubandhu calls on four alleged counterexamples in his attack on the continuity theory. They may be presented as follows:

(1) Consider a continuous, extended whole *w*, which is traversed by a moving object *x*. If *w* is simple, then *x* can at no time have traversed

only part of *w*; hence, *x* must traverse *w* not gradually, but in an instant, which is absurd.

(2) Consider a whole *w* and an observer *o*, who cannot see all of *w*. *O* sees one part of *w*, but not another; hence *w* has parts. To say that he sees *w*, but does not see *w*, at one and the same time, is a contradiction.

(3) Consider a continuous field *f*, with various elephants *e*, horses *h*, and cattle *c* grazing upon it. If *f* is simple, then *e*, *c*, and *h* are all in the same place, which is again an absurdity.

(4) Consider a little fish f_1 and a big fish f_2, which are similar in all respects, but size. An observer stationed an appropriate distance from f_1 and f_2 will be able to discern f_2, but not f_1. But how is this possible if they have just the same characteristics and the same number of parts?

Are these counterexamples really damaging to the continuity theory? If any one of them is, that will be quite enough for Vasubandhu's purposes. Each requires more careful consideration. In particular, it will be necessary to weigh them against recent Aristotelian suggestions put forth by R. M. Chisholm, e.g., that our "part of" locutions may be replaced by clauses in which "partly" is used as a predicate modifier.[18]

(1) There can be no doubt that a sentence such as "*x* has traversed part of *w*" can be rendered, e.g., "*w* has been partly traversed by *x*." But what of, "By now *x* has traversed a larger part of *w* than it had a minute ago"? We might try something like:

(i) By now *w* has been partly traversed by *x* ; (ii) a minute ago *w* had been partly traversed by *x* ; (iii) *w* is possibly such that it is divided into *either F* and *G*, *or H* and *I*, but not into *both F* and *G and H* and *I*; (iv) *F* is (tenselessly) smaller than *H*; (v) if *w* had been divided into *F* and *G* a minute ago, *x* would have traversed *F*; and (vi) were *w* divided into *H* and *I* now, *x* would have just traversed *H*.

The atomist, who can simply say, "By now *x* has traversed numerically more minimal parts of *w* than it had a minute ago," does have the advantage of relative simplicity of expression in this instance. Nor is our example, by any means, a terrifically complex one: the mind boggles to think how an Aristotelian might describe the progress of a number of objects which were

traversing the same object along various paths and at different speeds. This is not to say that Aristotle's theory must be false, but was it not advanced in the name of simplicity?

(2) The Aristotelian response to Vasubandhu's second counterexample might be: "*W* is partly seen by *o*, and partly not." Complexity becomes a problem again if we turn our attention to possible translations of fairly straightforward part expressions, such as: "*O* sees part of the red part of *w*, and that part which he sees is partly soiled."

(3) The challenge here is to banish reference to parts from such sentences as: "*E* stands on one part of *f*, *h* on another part, *c* on another, and there are parts of *f* intervening between, and surrounding, the parts on which *e*, *h*, and *c* are standing; and on those parts there is nothing at all."

Intricate juggling of the "possibly such that it is divided" locution may enable the Aristotelian to accomplish the appropriate translations in (2) and (3)—I leave it for the reader to experiment—without loss in meaning. Certainly, a great harvest of convoluted expression may be reaped in this fashion.

(4) Vasubandhu's final counterexample is quite strange, and I am not at all certain that I have grasped his meaning.[19] The point seems to be that difference in size between similar objects is explicable in realist terms only by either the assumption that the objects have different quantities of minimal parts, or the assumption that relations among their respective parts somehow differ. But if things have no parts, there can be no such differences. If this is so, then distinctions of size become unintelligible. Now, if there is any truth to this, it will be somewhat more troublesome for the Aristotelian than counterexamples (1)-(3); for, as we have already seen, any spatially extended whole, of whatever size, has precisely an infinite number of potential parts. Thus, the doctrine of possible division cannot be meaningfully invoked here; that is, if the number of, or relations among, minimal parts are fundamental to difference of size, then only actual parts will do. The intuition is, I think, clarified if we consider a possible universe, along lines suggested by Max Black, containing only two spheres similar in every respect except size: both have infinite possible parts, and the relations among these parts are in every aspect similar. Of course, one might posit absolute space, and go on to ground size difference in the relations that hold between possible parts and points in absolute space, but this would necessitate possible parts having actual relations, and would require the supposition of real atomic points as well. So it is not clear that we have thus gained any ground.

"But is not, say, 'bigger than' a primitive relationship? Is it not simply the case that the one sphere is bigger than the other?"

If "seeming bigger than" is what is meant here, then of course it is a primitive relationship holding between phenomenal objects—even the idealist can agree to that! But phenomenal relationships, in the absence of other considerations, imply nothing ontologically; they merely tell us something about how things appear to perceivers. On the other hand, it is not nearly so clear that "*being* bigger than" is, in fact, a *primitive* relationship. Imagine that the aforementioned two spheres are in far distant corners of the universe. In the region possessing the smaller sphere, matter is more dense, gravitational conditions are different, and so the people are "smaller" than in the region that has the larger sphere. Scientists from the two regions make radio contact and discuss their respective spheres. Incredibly they conclude that both spheres are two yards in diameter and weigh 1500 pounds.

"The fact that the scientists cannot *know* that there is a size difference hardly goes to show that there *is not* one."

True enough, but minimal part atomism does permit knowledge, too, in this instance. If atoms are similar throughout the universe, then eventually the scientists will discover differences between the spheres and realize that they had been mistaken up until that time. And is not a science that corresponds to *being* preferable to one that does not?

At this writing, I am not at all certain that Vasubandhu's four counterexamples can be made to do the work demanded of them, i.e., demonstrate that, on the assumption that matter is real, atomism necessarily follows. They do, however, underscore several of the difficulties which must be resolved by the continuity theory, if it is to be made capable of fully expressing our commonplace intuitions with respect to parts and wholes.

Vasubandhu sought to prove idealism by demonstrating, like Kant, that our concept of composite material wholes necessarily entails there being simple, atomic substances and that atomism is false. In so doing, he, no less than the latter, underestimated the potential strengths of the continuity theory and the point-particle theory. It may be asked: Would even a fully successful argument of this type be sufficient to demonstrate the truth of idealism? It is of interest to remark that, among Vasubandhu's successors in the Buddhist tradition, few seem to have thought so: Dharmakīrti, for example, though probably leaning in favor of some form of idealism, saw in Vasubandhu's mereological argument at best a supplement to profounder epistemological arguments for idealism;[20] and Śāntarakṣita, who refined Vasargument to a greater degree than did any other classical Indian thinker, saw here not a "proof of idealism," but a step on the way to a near total rejection of speculative metaphysics.[21] Still, Vasubandhu must be cred-

ited with making significant progress in the development of Indian mereological thought and with constructing not a few puzzles, for which decisive answers are wanting even today.

I have endeavored here to provide a clear and critical account of Vasubandhu's chief argument in support of an idealist ontology, an argument that represents but one aspect of a richly elaborated philosophical system, which merits more thorough *philosophical* interpretation overall than it has received to date. It is my own conviction that the value of such inquiry derives not from the discovery that, for instance, the ancient Indian philosopher in question was in his musings particularly right or wrong. What is of interest is that he mused on these matters at all. When we begin to appreciate Vasubandhu's insights from the vantage point of our own philosophical understanding, what is most human about us leaps through centuries, rushes across continents, and greets what is most human in what had formerly been other. We meet Vasubandhu face-to-face, incline toward one another, and commune in our perennial capacity to puzzle over what is real.

Appendix: *Viṃśatikā*, 11-15;
The Proof of Idealism[22]

i. *Rejection of possible alternatives to idealism* (v.11).

How, moreover, can it be known that the Transcendent Lord spoke of the existence of the domain of form with this intention,[23] and that no object of any particular perception of form, etcetera,[24] does, in fact, exist?

v.11 **The object is not simple,**
Nor is it atomically complex,
Neither is it an aggregation;
For the atom cannot be proven.

What does this assert? Whatever, in the domain of form, etcetera, may be an object of a particular perception of form, etcetera, is possibly either (*i*) a part-possessing form, as conceived by the Vaiśeṣikas; or (*ii*) atomically complex; or (*iii*) an aggregation of the atoms together. But (*i*) the object is not simple, because there is no apprehension whatever of a part-possessing form discrete from the parts; (*ii*) neither is it complex, because single atoms are not apprehended; and (*iii*) their aggregation is not the object, for the atom *qua* simple substance cannot be proven.[25]

ii. *Refutation of atomism* (v.12-14).

ii.a. *Refutation of the atomist supposition of the conjunction of partless minimal parts* (v.12-13).

ii.a1. *A dilemma for one who affirms the conjunction of such minimal parts* (v.12).

Why can it not be proven? Because:

v.12ab **Owing to the simultaneous conjunction of six,**
The atom must be six-parted.

If there be simultaneous conjunction with six atoms from six directions, then the atom must be six-parted; for in the place of one there cannot be another.[26]

v.12cd **If the six occupy the same place,**
Then the cluster must be a mere atom.

Now, let the place of one atom be that of six. In that case, because all occupy the same place, the entire cluster must be a mere atom; for they do not exclude one another. Thus, no cluster whatever is found.

ii.a2. *A dilemma for those who posit the conjunction of aggregations* (v.13).

The Kāśmīri Vaibhāṣikas, to avoid the fault implicit in partlessness, namely, that the atoms cannot conjoin, maintain that the aggregations conjoin with one another. Let us demand of them whether or not the aggregations of atoms be objects discrete from those (atoms).

v.13ab **If not a conjunction of atoms,**
Their aggregation is that of what?

"That" here stands for "a conjunction."

v.13cd **It is not merely owing to partlessness**
That their conjunction cannot be proven!

Now, should it be that the aggregations, too, do not conjoin with one

another, then let it not be said that partlessness is the reason that the conjunction of atoms cannot be proven; for the conjunction of aggregations which have parts cannot be affirmed.

ii.a3. *Conclusion of* ii.a.

Therefore, it cannot be proven that the atom is a simple substance.

ii.b. *The incompatibility of spatial extension and simplicity* (v.14).

Whether the conjunction of atoms be affirmed or denied:

v.14ab **That which is spatially extended**
Cannot be simple.

If the atom be spatially extended—that is, e.g., if it has a front part that is different from its bottom part—how can an atom of such nature be simple?

v.14c **How can there be shade and shadow?**

If not a single atom is spatially extended, then, when the sun is shining, how is it that in one place there is shade, while there is sunlight in another? (Under such circumstances, the atom) does not, after all, have other parts that are *not* struck by the sunlight. And how can one atom be in the shadow of another, if spatial extension be denied? Moreover, if no atom whatever has discrete parts, such as might obstruct passage from one place to another, then, in the absence of such obstruction, owing to the cospatiality of all, an entire cluster must be a mere atom, as has already been asserted.

But do you deny that shade and shadow are of the cluster, rather than of the atom?

Are you, then, maintaining them to be of a cluster which is discrete from the atoms? That cannot be, for:

v.14d **If the cluster be not discrete,**
they are not of it.

If the cluster is not held to be discrete from the atoms, then it is established that they (shade and shadow) are not of it (the cluster).

iii. *Refutation of the doctrine of simple, extended wholes* (v.15).

So long as the defining characteristic of form, etcetera,[27] is not negated, does it matter whether or not the constructed idea (of a particular form corresponds to) atom or aggregation?[28]

What, then, is their defining characteristic?

"Being an object of the eye, etcetera" and "being blue, etcetera."[29]

This must be examined: Is that blue, yellow, or other, which is held to be the object of the eye, etcetera, a simple substance, or is it complex?

What are you driving at?

The fault of complexity has already been stated, and:

v.15 **If it be simple, there cannot be gradual motion;**
Neither is simultaneous apprehension and nonapprehension possible;
Neither can there be separate, varied presence
Nor can there be unseen minuteness.

If one imagines that, so long as the visual object (blue, yellow, etc.) is unbroken, it is not many, but is one substance, then gradual traversal of the earth becomes impossible, "traversal of" meaning "motion across"; for a single step would have to traverse the whole.

Neither could there be apprehension of the near part (of some object) simultaneous with nonapprehension of the far part. At that time, its apprehension and nonapprehension is impossible.[30]

Neither could there be the presence in a single place[31] of separate and varied elephants, horses, and so forth; for, if one be in the location of another, how can they be held to be separate? How is it that (the places) they occupy and do not occupy are one, when empty space is apprehended between them?

Moreover, minute aquatic creatures having the same features as large ones would not be imperceptible if discreteness of substance were known only through characteristic difference, and nothing else.[32]

iv. *Conclusion*

Therefore, atomic distinctions must be supposed and that (atom) is not proven to be simple. It being unproven, it is not proven that forms, etcetera, are the objects of the eye, etcetera. Thus, it is proven that they are mere perceptions.

The University of Chicago

Notes

[1]T. W. Rhys Davids (trans.), *The Questions of King Milinda*, pp. 43-44, in F. Max Müller (ed.), *The Sacred Books of the East*, Vol. XXXV (Oxford, 1890). The Graeco-Bactrian king Menander (Pali: Milinda) reigned during the latter part of the second century B.C. His dialogues with Nāgasena were not set down until a somewhat later date and have little historical value. In particular, they offer no dependable evidence of possible philosophical exchange between Greece and India. Arguments based on the dissection of a thing into its constituent parts were well known in India from at least the first half of the first millenium B.C. Cf. Steven Collins, *Selfless Persons* (Cambridge: Cambridge University Press, 1982), pp. 78-84, and *passim*. Several centuries prior to the age of Menander, the elements of the person had already been metaphorically compared to the parts of a chariot, including horses and driver, in *Kaṭhopaniṣad*, III.3ff., and in many of the early Buddhist *suttas*.

[2]Jonathan Barnes, *The Presocratic Philosophers* (London, 1979), vol. I, p. ix.

[3]This is not to say that philology has in fact already completed its work in this case. On the contrary, we still face the vexing problem of whether the name Vasubandhu was that of one philosopher or of several. Nonetheless, the text studied in this essay is unquestionably well enough established for present purposes.

[4]I am thinking chiefly of the brilliant line of logicians beginning with his disciple Dignāga.

[5]Quotations from Vasubandhu given below are based on my own translation, which is appended. I have, however, allowed myself some liberties in the body of the paper to avoid the more stilted style of the appendix, reference to which is given by means of the lowercase Roman numerals used in the section headings.

[6]Aristotle, *On Generation and Corruption*, Bk. I, Ch. II, 316a. 25-30. Cf., also, Fred D. Miller, Jr., "Aristotle Against the Atomists," in Norman Kretzmann (ed.), *Infinity and Continuity in Ancient and Medieval Thought* (Ithaca: Cornell University Press, 1982), pp. 87-111.

[7]Vasubandhu, unlike many of his successors, seldom gives clear syllogistic expression to his arguments. Hence, reconstructions like this one may only be given with some reservations. The conclusion (section iv.) of Vasubandhu's discussion, however, comes very close to setting out the present syllogism in unambiguous terms.

[8]The best discussion of Nyāya-Vaiśeṣika mereology to date will be found in Karl H. Potter, *Encyclopedia of Indian Philosophies; Indian Metaphysics and Epistemology: The Tradition of Nyāya-Vaiśeṣika up to Gaṅgeśa* (Princeton, 1977), pp. 74-86, and passim.

[9]One question not raised here is: Even if the whole is not a substance in addition to the unified collection of parts, is it not still true that the whole is something more than the sum of its parts? I see no objection, from Vasubandhu's standpoint, to the notion that a unified collection might have properties which neither its individual parts, nor any mere collection of those parts, might have. Strong conjunction is, in fact, just one such property.

[10]The considerations mentioned in n. 7 apply here as well.

[11]I think it at least questionable whether Vasubandhu ever did, in fact, seriously entertain the notion of unextended material points. It is referred to explicitly in none of his works, so far as I know. His commentators who do mention it, e.g., Saṃghabhadra and Yaśomitra, may have posited such points in an attempt to formulate a realist response to Vasubandhu's idealism. See below, n. 13, as well as section ii.b., v.14c.

[12]Y. Karunadasa, *Buddhist Analysis of Matter* (Colombo, 1967), pp. 12-14, especially p. 13: "Where there is one *sapratigha* [resistant] object there cannot be (at the same time) another." Karunadasa here bases his statement on works of Vasubandhu and his commentators.

[13]In my discussion, I interpret Vasubandhu *as if* he were addressing himself to a fully developed form of the theory in which atoms are supposed to be unextended. As has already been noted, however, *that* form of the theory may well have postdated Vasubandhu. Here, his concern was probably an earlier version which postulated conjunction, not among atoms, but among clusters of unconjoined, simple, but extended atoms.

[14]Cf. C. D. Broad, *Kant—An Introduction* (Cambridge, 1978), p. 224: "I do not see the slightest objection to the view that every body occupies the region which it does occupy *only discontinuously*." Similarly, P. F. Strawson, *The Bounds of Sense* (London, 1966), p. 184: "there would be no contradiction or absurdity in a physical theory according to which a composite material body was made up of a finite number of simple and unextended point-particles each of which was the sole occupant of a part of the space occupied by the material body as a whole." Strawson's theory, however, allows for the formation of unified collections; for each particle might "exert some causal power throughout the region of space of which it was the unique occupant."

[15]Cf., e.g., Jonathan Barnes, *The Presocratic Philosophers*, vol. I, ch. XII; Aristotle, *Physics*, 231b.15; and the antithesis of Kant's second antinomy.

[16]Compare Pierre Bayle: "every extension, no matter how small it may be, has a right and a left side" (quoted in James Van Cleve, "Reflections on Kant's Second Antinomy," *Synthese*, 47 (1981), p. 483).

[17]A. L. Herman, in his recent *An Introduction to Buddhist Thought* (Lanham, MD: University Press of America, 1983), pp. 336-40, has attempted an analysis of Vasubandhu's attack on atomism which, I believe, is flawed. Rather than undertake a point by point critique of Herman's discussion, it will be sufficient to regard only his proposed counterargument addressed to Vasubandhu's argument from shadow and shade, for it seems to me that most of Herman's difficulties are revealed by a consideration of this one short passage:

> The realist defender of atoms, whoever he may be, is in a good position to attack the apparently devastating dilemma of atomism with a plausible and sensible counter-dilemma of his own. Let's call it "counter-dilemma of atomism":
>
> 1. If the atom (or an aggregate) has spatial divisions then there would be shadows, etc.
> 2. If the atom (or an aggregate) has no spatial divisions then it can be a unity.
> 3. But the atom (or an aggregate) either has spatial divisions or it has no spatial divisions.
> 4. Therefore, either there are shadows, etc., or there is a unity, either of which is perfectly sensible.
> 5. Therefore, the atomist's position is perfectly sensible and not absurd.
>
> As things stand, who is to say which dilemma or counter-dilemma, Vasubandhu's or the atomist's, is the most defensible? Thus the atomist's possible reply to Vasubandhu.

It should be clear that Herman has got himself into a muddle here by, among other things, completely failing to understand Vasubandhu's first dilemma of conjunction (ii.a1. above); for his third premise asserts precisely what Vasubandhu there refuted. Neither has he, in

speaking of "a unity," drawn any distinction whatsoever between the unity of a continuous whole, that of an extended minimal part, and that of a point-particle, which distinction, as I indicated above, does in fact reveal the weakness in the argument from shadow. This being so, it is altogether unclear just how his fourth step provides us with sensible alternatives. I ignore here several purely formal flaws in the argument, which require only relatively simple reformulation.

[18]On this, see J. Van Cleve, "Reflections on Kant's Second Antinomy," p. 490.

[19]A somewhat different interpretation may be found in C. Hamilton, *Wei Shih Er Shih Lun* (New Haven, 1938), p. 57, n. 103: "On the hypothesis of unity, each occupying subject must fill the whole of the place occupied. Hence in a single body of water each animalcule would have to be equal to the whole. Consequently small and large would be of the same measure—an absurdity."

[20]Dharmakīrti, *Pramāṇavārttika*, ed. Svāmī Dvārikādāsaśāstrī, Bauddha Bhāratī Series (Vārāṇasī, 1968).

[21]Śāntarakṣita develops the argument at great length in his "Examination of External Objects" (*Bahirarthaparīkṣā*), in his *Tattvasaṅgrahaḥ*, vol. 2, ch. 23, Bauddha Bhāratī Series 2 *(Vārāṇasī, 1968). He applies it to the general critique of substantialist theories in his Ornament of the Middle Way* (*Mādhyamakālaṃkāra*), which, with its commentaries, survives now only in Tibetan translation (Tohoku 3884ff.).

[22]In translating this passage from the Sanskrit, with reference to the eighth century Tibetan translation, I have utilized the following texts and earlier translations: L. de la Vallée Poussin, "Viṃśakakārikāprakaraṇa," in *Le Muśeon* (Louvain, 1912), pp. 53-90; S. Lévi, *Vijñaptimātratāsiddhī* (Paris, 1925); Idem., *Materiaux pour l'étude du système Vijñaptimātra* (Paris, 1932); Clarence H. Hamilton, *Wei Shih Er Shih Lun* (American Oriental Series, vol. 13, New Haven, 1938); N. Aiyaswami Sastrin, *Viṃśatikā* (Gangtok, 1964), K. N. Chatterjee, *Vijñaptimātratāsiddhi* (Varanasi: Kishor Vidya Niketan, 1980); Mrinalakanti Gangopadhyaya, *Indian Atomism* (Atlantic Highlands, NJ: Humanities Press, 1981).

[23]The Buddhist idealist must argue that the references to material objects made by the Buddha in his discourses are intended not as metaphysical declarations concerning the real order of things, but merely as popular expressions, conforming to philosophically uninformed convention.

[24]"Form, etcetera" = "visual objects and the objects of the remaining four physical senses."

[25]"Cannot be proven" (*na sidhyati*) = "cannot be shown to exist," or "is an inadmissible hypothesis."

[26]*ekasya yo deśaḥ tatrânyasyâsaṃbhavāt*, cf. n. 12 above.

[27]See above, n. 24.

[28]*Sannivésaparikalpa eṣaḥ*, the Sanskrit phrase with which this passage opens, rendered *here as "the constructed idea," is somewhat obscure, as is its Tibetan translation at this point. See Hamilton's note on it (Wei Shih Er Shih Lun*, p. 53, n. 93), and Sastrin's remarks based on the gloss of Vinītadeva (*Viṃśatikā*, pp. 47-48, n. 50).

[29]This is, of course, an abbreviated, enumerative definition: physical things are considered by Buddhist realism to be possibly such that they are objects of the five senses, or possibly such that they instantiate the properties that are perceived by those senses. See, e.g.,

Karunadasa, *Buddhist Analysis of Matter*, ch. I.

[30]*na hi tasyâiva tadānīṃ grahaṇañ câgrahaṇañ ca yuktam.*

[31]The Sanskrit text reads *anekatra* "here and there," but this is to be emended to *ekatra* "in a single place"; cf. the Tibetan *gcig-na*, "in one place."

[32]Compare Hamilton's translation of Hsüan Tsang's Chinese version: "Moreover there should also be no such scarcely perceptible tiny things as water animalcules, because being in the same single space with the coarse things they should be of equal measure," *Wei Shih Er Shih Lun*, p. 57.

Bimal Krishna Matilal **Causality in the Nyāya-Vaiśeṣika school**

Causality implies a host of philosophic problems which are pertinent to different branches of philosophical studies. My aim here is, however, a modest one. While explaining the notion of cause as it was understood by different philosophical schools of classical India, I shall try to show that (a) concern for the notion of causation was taken more seriously in some Indian philosophical schools than it had been in some of their Western counterparts, and (b) that the meaning of 'cause' (*kāraṇa*) is much wider in Indian philosophy than it is in the West.

To substantiate the second point first, I shall first very briefly refer to the classification of 'causal conditions' as found in the *Abhidharmakośa* of Vasubandhu and then to another classification found in the Vaiśeṣika school. Two different terms, *hetu* and *pratyaya*, are used to cover all causal notions in the Abhidharma system. There is, however, no essential difference between these two concepts, the six types of *hetu* being included under the broader notion of *pratyaya*.

The *Abhidharmakośa* first explains six types of *hetu*, of which the first is called the *kāraṇa-hetu*, the 'prototype' of a *hetu*. The definition of a *kāraṇa-hetu* reflects the attitude of the Abhidharma school toward the general notion of cause. In fact, to be a cause in this sense means to exist without being a 'hindrance' (*vighna*) to the production of the effect in question. Thus, if *x* is a *dharma* that originates, then any *dharma* other than the 'own-being' (*svabhāva*) of it may be called its *kāraṇa-hetu*.[1]

Four types of *pratyaya* are: (1) *hetu-pratyaya*, (2) *ālambana-pratyaya*, (3) *samanantara-pratyaya*, and (4) *adhipati-pratyaya*. Of these, the first includes the 'remaining' five *hetus* in the foregoing classification while the last includes the 'prototype', that is, the *kāraṇa-hetu*.[2] The second and the third are found useful only with regard to the states of consciousness. The *ālambana* is the 'causal basis', that is, the object, of a particular state of consciousness which is regarded as an effect or event. The *samanantara* is the 'causal precedent', that is, the preceding moment of consciousness conditioning the succeeding moment. It should also be mentioned that in the remaining five *hetus* (a discussion of which I will skip here) of the six types mentioned earlier are included *hetus* like a *sahabhū hetu*, where two *dharmas* originating simultaneously are said to mutually condition each other.

This breathtaking classification of the *Abhidharma* will be enough to show *hetu* or 'cause' is being used here in its widest possible sense. It is sometimes pointed out that we should not probably translate *hetu* or *kāraṇa* in this context as a 'cause' or a 'causal condition'. But I am in favor of retaining such translations simply because such terms as '*hetu*' and '*pratyaya*' mean almost the same thing (and, by the same token, share the same ambiguity) in ordinary Sanskrit

Bimal Krishna Matilal is Professor of Sanskrit Studies at the University of Toronto and is Editor of the Journal of Indian Philosophy.

as terms like 'cause' and 'causal condition' do in ordinary English. It is a philosopher's privilege to use an ordinary term in a highly specialized sense.

If to cause an effect in the *Abhidharma* means not to obstruct the production of that effect (that is, that *dharma*), to cause an effect in the Vaiśeṣika system means to be its immediate (and 'unconditional') antecedent. The Vaiśeṣikas speak of a causal substrate or a 'substantial' cause (*samavāyikāraṇa*) which is the substance where the effect occurs (through the relation of inherence). Ordinarily, this is the material cause of an effect, even comparable to the 'material' cause in the fourfold division of Aristotle. But strictly speaking, this is a wider notion in the Vaiśeṣika system, since such 'nonmaterial' substances as the soul and the physical space are said to be causal substrates of suitable effects. For example, a state of consciousness as an event is said to occur in the soul (that is, the person) which is its causal substrate.

All causal conditions other than the causal substrate that are relevant to the effect are classified by the Vaiśeṣikas into two groups: *asamavāyin* cause (nonsubstantial, literally 'noninherent' cause), and the *nimitta* 'efficient' cause. The former forms an artificial group which includes only qualities (and relational qualities like 'conjunction' *saṃyoga*) that *inhere* in the causal substrate and are causally relevant to the effect.[3] For example, color of the threads causes the color of the cloth, or conjunction of different parts of a table causes the table in this manner. The group of 'efficient' cause includes everything else that is causally relevant to the effect in question. Thus, the potter, along with the potter's wheel, rod, water, etc., is called the 'efficient' cause of the effect, pot. Aristotle's notion of the 'efficient' cause, in a liberal interpretation, can match this Vaiśeṣika notion. But nothing like Aristotle's notion of the 'final' cause or the 'formal' cause can be found in the Indian schools.[4]

The distinction of a 'nonsubstantial' cause from the group of efficient causes seems to be artificial. It was probably based upon the awareness that as long as the effect, the color of a cloth, exists its 'nonsubstantial' cause (that is, the color of the threads) should, like the causal substrate, also exist. But an 'efficient' cause like the potter's wheel may wither away (after the effect is born) without affecting the effect in any way. In other words, the father may die after the son is born, but what inheres in the son's body and limbs must stay as long as the son is alive.[5] A notion parallel to that of a 'nonsubstantial' cause of the Vaiśeṣikas is hard to find in the Western tradition.

Concern for a causal notion was fundamental in almost all schools of Indian philosophy. We should remember that philosophic activity in India arose out of the cosmogonic speculations of the Vedas and the Upaniṣads. The all important business of philosophy was to attempt to discover some simple, unitary cause for the origin of this complex universe. Various alternative theories were propounded from the very early period, as is well evidenced in the passages of the Śvetāśvatara Upaniṣad, and in the *Nyāyasūtra*, chapter 4.[6] Some of the main views about the original cause of the universe were: (a) time,

(b) nature or 'own-nature', (c) *pradhāna* 'the unmanifest' matter, (d) god, (e) atoms, and (f) cyclical dependent origination. Of these, I shall briefly allude to the three important views, the Sāṃkhya, the Buddhist, and the Nyāya-Vaiśeṣika views.

The earliest critique of the notion of cause is to be found in India in the *satkāryavāda* doctrine of the Sāṃkhya school, which was historically the earliest of Indian philosophic systems. The doctrine means that the so-called effect preexists in its cause, causation being merely a change of transformation from one state to another while the original 'thing' (cf. *dharmin*) remains constant and unchanging. An effect means a change in only the attributes or characteristics of the thing, a new state of affairs means manifestation of what was *potentially* present (cf., *sat*) in the early state of affairs, that is, in its so-called cause.[7] The Sāṃkhya metaphysics posits an ultimate, original matter, the *pradhāna* or *avyakta* (the 'chief' or the 'unmanifest') from which the whole material world evolves. It is also admitted that the potential becomes actual at every moment and thus transformation of the world is automatic and instantaneous. This aspect of the Sāṃkhya theory might have influenced the Buddhist doctrine of universal flux.

The Buddha's doctrine of impermanence (*anityatā*) was developed by the later Buddhists as the doctrine of universal flux. Causation in Buddhism is spanned by its doctrine of conditioned (or dependent) origination (*pratītya-samutpāda*). In Abhidharma, for example, every *saṃskṛta dharma*, that is, every effect, is said to originate merely depending upon what is called its *pratyaya* (or *hetu*). Like the empiricist in the West, the Buddhists reduced causal dependence to nothing more than a mere sequence, a very loose relation usually expressed in the formula: if *x* is, *y* arises (cf., *asmin sati idaṃ bhavati hrasve sati dīrghaṃ yathā*). The Mādhyamikas rejected the notion of *savabhāva* 'own-being' of things as well as origin (*udaya*) and decay (*vyaya*) of things (*bhāvas*).[8] Thus, they might have paved the way for the doctrine of non-origination (*ajāti-vāda*) of Gauḍapāda.[9] The popular belief that an effect is brought about by its causes is reduced to an absurdity by the Mādhyamika dialecticians.

The doctrine of universal flux is the result of the Buddhist rejection of the notion of *potentiality*. To be causally potent means here only actual production of the effect. Everything is in a flux, there being so stability (*sthiratva*). Causation is simply incessant succession of events. Unlike the Sāṃkhya, the Buddhists rejected the notion of any unchanging core of things, any unchanging *dharmin* underlying the seeming fluctuations.[10]

The Nyāya-Vaiśeṣika school opposed the Sāṃkhya by its doctrine of *a-sat-kārya-vāda*, according to which an effect is a new creation, and hence numerically different, from its cause. The description of causation in this school is closer to the commonsense view of a cause. True to the spirit of empiricism, causal relation is described here also as one of invariable sequence. Udayana

asserted, in answer to the skepticism about causality, that a causal explanation of an event is needed unless we want to settle for a total accidentality or 'whimsicality' of everything. A particular effect happens at a particular moment, not always. This is what is called the 'temporality' (*kādācitkatva*) of an effect, and this temporality implies dependence of the effect upon something other than itself. Causal relation is nothing more than this obligatory dependence.[11]

The Nyāya school rejected also the notion of *śakti*, 'power', 'efficacy' or 'force', connected with causation. The Mīmāṃsakas, being consistent with the common belief, argued in favor of an efficacious power or *śakti* present in the cause to produce the effect—a power which can be destroyed by the presence of an 'antidote' (*pratibandhaka*) and can conceivably be resuscitated by an 'antidote to the antidote' (cf., *uttejaka*). Thus, fire burns because of its power to burn—a power which either can be destroyed or resuscitated under suitable conditions. While refuting the notion of *śakti*, Udayana solved the problem presented by the influence of antidotes by boldly asserting that causation implies presence of not only 'positive' causal conditions but also of relevant 'negative' conditions. For a particular effect to happen, the *absence* of the relevant antidote is also needed as one of its causal conditions.[12]

The Navyanyāya treatment of causation is interesting in many ways. The notion of invariable sequence is explained as holding between generalities rather than between particular events. Thus, Śaśadhara defined a cause as one belonging to a class, individual members of which invariably precede individual members of another class, the relevant effect-class (cf., *kāryaniyatapūrvavṛttijātīyatva*).[13] A potter's wheel is said to be a cause of a particular pot because it belongs to the class of those wheels, members of which are *seen* to precede invariably the production of members of the pot-class. The notion of invariable sequence is, however, to be derived from experience, from what is called *anvaya* and *vyatireka* (seeing cases of association and absence).

Although in rejecting the notion of 'efficacy' or 'power' Nyāya resembled the position of David Hume, the doctrine of invariable sequence was not propounded here exactly in Humean spirit. For Hume, it is only the mind that spreads itself on external objects and conjoins them as cause and effect while nothing really exists between them to be so conjoined.[14] This is more like the Buddhist view than the Nyāya view. For Nyāya, invariable sequence is discovered by the mind but it exists between extramental realities like universals or class characters. Perhaps Nyāya shuns the Humean empiricism while it asserts its doctrine of real universals. In fact, although the early Nyāya-Vaiśeṣika doctrine of universal was modified by Navyanyāya, it still maintained that certain class-characters were real in order to explain, among other things, the relation of cause and effect.[15]

Thus, it was felt in Navyanyāya that the criterion of invariable sequence was not enough for distinguishing causal conditions of a particular effect. For example, the production of a pot is preceded invariably not simply by the

potter's wheel but also the color and circularity of the potter's wheel. But the color of the wheel is immaterial and irrelevant to the production of the pot. To exclude such irrelevant items from being considered as causes of the particular effect in question, such Navyanyāya authors as Śaśadhara and Gaṅgeśa introduced the notion of what they called *ananyathāsiddhatva* 'unconditionality'. The invariable presence of the color of the wheel before the pot is produced is *conditioned* by the presence of the wheel itself, and hence it need not be taken into account while we consider the relevant (causal) conditions for the effect in question. By same token Navyanyāya excludes the cause of a cause from being considered as a cause of a particular effect. This takes care also of a conceivable case where each time a pot is produced on the potter's wheel, a donkey always walks by immediately before the event. Experience of invariable sequence may demand that we construe the two events as causally related. But Nyāya claims that the 'unconditionality' criterion can save the situation since it is possible to find a reasonable explanation of each case of such appearances of a donkey. If such reasonable explanation is found, the donkey's appearance will no longer be an 'unconditional' antecedent.[16]

If the Navyanyāya analysis of the causal relation seems to be somewhat embarrassing to a Humean empiricist, we may introduce here a brief discussion of some contemporary problems of causation. David Hume's critique of causality brought into focus two important questions in contemporary philosophy. The first question is: whether there is, after all, any necessary connection between a cause and its effect? The second, and perhaps more basic question is: whether the notion of cause is at all a viable concept in philosophy, useful for explanation and understanding of what are called 'events'?

Because of the muddle and complexity associated with the notion of 'cause', some modern philosophers (notably Bertrand Russell) despaired of making any sense of the word 'cause', and hence recommended "its complete extrusion from the philosophical vocabulary."[17] The rise of modern theoretical physics, some doctrines of which throw doubt even upon the once universally acclaimed *universality principle of causation* (the principle that states, "every event must have a cause"), has contributed further to the modern despair about the notion of cause. But whatever may be the situation in theoretical and higher physics, it is almost undeniable that the concept of a cause is quite useful in the common affairs of life, in applied technology, in moral fields, in law and jurisprudence. Thus, I believe a philosopher can hardly afford to be totally indifferent to this concept. It will be enough to point out here that respectable modern philosophers such as A. J. Ayer, R. G. Collingwood, C. J. Ducasse, and G. E. van Wright, have found the concept of causality to be useful.[18]

Modern despair about the notion of cause can partly be ascribed to Hume's rigid demand for the empirical analysis of causation. Hume used the notion of similarity or resemblance in giving an empirical explanation of cause. A cause is defined as "an object precedent and contiguous to another, and where all the

objects resembling the former are placed in a like relation of priority and contiguity to those objects, that resemble the latter."[19] Unfortunately the notion of similarity invites a number of philosophic problems in empiricism. For one thing, an event today must be similar to an event of yesterday in some *essential* respects in order to be caused by a similar cause. We may recall that Navyanyāya introduced the notion of real universals or class character. To avoid this quandary of empiricism, a follower of Hume might argue that similar effects, in order to have similar causes, must be similar only in certain *crucial* or *relevant* respects. It is, however, easy to see that the notion of such "relevance" cannot be explained without resorting to the causal notion, and thus it will involve us in a circularity.

The question of relevance may enter in the discussion of causal relation in another way. If, following Hume, the notion of necessity is completely eliminated and causation is reduced to mere invariable sequence, then, as Thomas Reid pointed out against David Hume, we will have to admit that day is the cause of night and vice versa, since day is seen to be invariably followed by night and night by day.[20] J. S. Mill referred to this criticism of Reid and defended Hume by introducing the notion of 'unconditionality' in defining the notion of cause. Mill argued as follows: "If there be any meaning which confessedly belongs to the term, it is *unconditionalness* . . . . The succession of day and night evidently is not necessary in this sense. It is conditional on the occurrence of other antecedents."[21] Since it is conceivable that there could be circumstances (for example, sudden stopping of the rotation of the earth) under which day will not be followed by night and vice versa, Mill said, the two are *not unconditionally* conjoined and hence not causally connected. It is unfortunate that Mill here abandoned, quite unconsciously, the very point of the empirical analysis of Hume, and reintroduced the notion of necessity disguised as 'unconditionality'. Besides, if, as Mill explained, conceivability of the circumstances under which a known invariable sequence may not follow is enough to upset the causal connection between events, then very few events could be said to be causally connected.

Mill's own explanation of the 'unconditionality' criterion was thus hardly satisfactory. However, one may justify the 'unconditionality' criterion in order to distinguish irrelevant antecedents from the causally relevant antecedents. But then we are thinking of what Navyanyāya calls *ananyathāsiddhatva* (which I have translated above as the criterion of 'unconditionality'), and not Mill's 'unconditionality' criterion.

The point at issue may be made in the following way. Suppose there is a unique tribe, each adult member of which has produced, after going through the usual tribal training period, a particular craft. This craft is unique to this tribe, and let us assume that no one else on earth has succeeded in producing that craft. Now suppose that each such adult member, after his training period and before producing the craft, has gotten up from bed at 6 A.M. on two succes-

sive Sundays. If invariable sequence is enough to establish causal connection, then not only the tribal training period but also their getting up at 6 A.M. on two successive Sundays should be considered causally relevant! In fact, to be true to the Humean spirit of empiricism, one has to accept such an apparently absurd consequence. The question of what is relevant, namely, the training period, and what is not relevant to the production of the craft cannot be decided unless some notion like the Navyanyāya idea of *ananyathāsiddhatva* is introduced. It may be noted that an Abhidharma Buddhist can nicely tackle the problem here by taking the bull by the horns. As noted earlier, he can claim that when something *has been* produced nothing that was prior to it was really irrelevant to its production. Vasubandhu exemplified the point as follows: When the village folk have successfully organized a feast without interference from the proverbial village troublemaker, they can say that the success of the feast was owing also to the help of that troublemaker. In other words, when an event is caused, almost everything, through its noninterference, can be causally responsible for it. Thus, the question of excluding irrelevant items from the domain of causation does not arise.

For Navyanyāya, the fact of their getting up at 6 A.M. on two Sunday mornings will be what is called *anyathāsiddha* 'conditioned otherwise', and hence it cannot be causally relevant. In other words, if we investigated each case separately, we could find in each case reasonable explanation of why that particular adult in that case got up at 6 A.M. on two successive Sunday mornings before the said craft was produced. And such an explanation would reveal that each of these facts was conditioned otherwise, *anyathāsiddha*. In the same way, Navyanyāya would declare that while being dry, that is, the lack of dampness, is a 'negative' causal condition for the matchstick to ignite, but the color of the matchstick is *anyathāsiddha*, being immaterial to its igniting.[22]

NOTES

1. See *Abhidharmakośa* and *Bhāṣya* of Vasubandhu, Part I, ed. Swami Dwarikadas Sastri (Varanasi: Bauddha Bharati, 1971), pp. 279–282.

2. Ibid., p. 349.

3. See Praśastapāda, *Padārthadharmasaṃgraha*, ed. Durgadhara Jha (Varanasi: Sanskrit Visvavidyalaya, 1963), pp. 244, 246.

4. This is a general comment, which may be subject to qualification. In fact, Asok Gangadean criticized this comment. But I am yet to be convinced by his arguments.

5. For this notion of *nimittakāraṇa*, see *Vyomavatī* of Vyomaśivācārya, commentary on *Praśastapādabhāṣya*, ed. Gopinath Kaviraj and Dhundhiraj Sastri, (Banaras: Chowkhamba, 1930), p. 140–142.

6. See *Nyāyasūtra* and *Nyāyabhāṣya*, ed. Ganganath Jha, (Poona: Oriental Book Agency, 1939), p. 246–270.

7. See *Sāṃkhyakārikā* and *Yuktidīpikā*, ed. Ramasamkar Tripathi, (Varanasi: B. Tripathi, 1970), p. 59–69.

8. See *Mādhyamikakārikā* of Nāgārjuna, ed. P. L. Vaidya (Darbhanga: Mithila Institute, 1960), chapts. 1 and 15.

9. See *Āgamaśāstra* of Gauḍapāda, ed. Vidhusekhara Bhattacharya (Calcutta: University of Calcutta, 1943), pp. 101–112.

10. For the Buddhist criticism of potentiality (*sāmarthya*), see *Ratnakīrti-nibandhāvalī*, ed. Anantalal Thakur (Patna: K. P. Jayaswal Research Institute) pp. 62–76.

11. See *Nyāyakusumāñjali* of Udayana, eds. P. Upadhyaya and D. Sastri, (Varanasi: Chowkhamba, 1957), pp. 41–60.

12. Ibid., pp. 103–117.

13. See *Nyāyasiddhāntadīpa*, ed. V. P. Dwivedi and Dhundhiraj Shastri, (Benares Cantt.: Lazarus & Co., 1924), pp. 89–92. A critical edition of this text has already been prepared by me and is forthcoming in the L. D. Indological Series, Ahmedabad.

14. David Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge (Oxford: The Clarendon Press, 1951), Book I, Part III, Sec. XIV, pp. 166–172.

15. See my *Epistemology, Logic and Grammar in Indian Philosophical Analysis*, (The Hague: Mouton, 1971), pp. 71–77.

16. See for a discussion of *ananyathāsiddhatva*, *Tattvacintāmaṇi* of Gaṅgeśa, ed. Kamakhyanatha Tarkavagisa (Calcutta: Asiatic Society of Bengal, 1884–1901), Part 2, vol. 3, pp. 154–155.

17. See B. Russell, *Mysticism and Logic* (1917; reprint ed., Garden City, N.Y.: Doubleday Anchor Books), pp. 174–201.

18. A. J. Ayer, *Foundations of Empirical Knowledge* (London: Macmillan & Co., 1951), Chap. 4. R. G. Collingwood, *An Essay on Metaphysics* (Oxford The Clarendon Press, 1940), Part 3-C. C. J. Ducasse, *Nature, Mind and Death*, Part II (La Salle, Illinois: The Open Court Publ. Co., 1951).

19. David Hume, *op. cit.*, p. 172.

20. Thomas Reid, *Essays on the Active Powers of the Human Mind*, Essay 4, included in *Works*, ed. Sir William Hamilton, (Edinburg: Maclachlan & Stewart, 1846–1863).

21. J. S. Mill, *A System of Logic*, 8th ed., Book III, Chap. V, Sec. 6 (1843; New York: Harper, 1881).

22. In fact, the *ananyathāsiddhatva* criterion cannot, in final analysis, successfully exclude all irrelevant factors from the relevant causal factor. Gaṅgeśa discusses the problem in *Tattvacintāmaṇi*, Part 2, vol. 3, p. 154–155. The important question, namely, why, of two invariably conjoined factors, one is taken to be the causal factor rather than the other, depends partly, according to Gaṅgeśa, on our intuitive notion of simplicity (*lāghava*) and relevance.

AN ONTOLOGY OF CONCRETE CONNECTORS[1]

As Nelson Goodman has rightly pointed out, the distinction between abstract and concrete is not the same as that between universal and particular (*The Structure of Appearance* (*SA*), pp. 200–201). Since, however, Goodman's definitions of these terms are rigged in the midst of his rather peculiar reconstruction, since those definitions therefore involve the use of the term "concretum" in the definiens, and since the term "concretum" involves a possibly dubious maneuver in its explication (cf. "A Query to Mr. Goodman" by the present writer, *Philosophical Studies,* December 1957, pp. 81–85), I prefer to formulate my own explication of "concrete" and "abstract."

The term "concrete" in its non-philosophical use carries with it many of the implications of the word "solid." "Abstract," on the other hand, means something from which an element has been left out in thought—or, which comes to essentially the same thing, something that has been separated from its (concrete) situation by thought. Roughly, then, the distinction is between that from which nothing has been removed in thought as over against that from which something *has* been removed in thought.

Now it is clear that this is not the same distinction as that between universal and particular. "Universality," as Goodman remarks, "is rather a matter of multiplicity of instances" (*SA,* pp. 201). Universals, if there are any, may be either abstract or concrete, which is to say that either an entity may be complete in all its details and still be repeatable (as Plato's Forms are usually understood to have been thought by him to be) or else that the repeatability of an entity always is due to the abstraction by men's minds from certain aspects of concrete situations.

Now returning to my explication of "abstract" and "concrete," it would seem that this distinction, unlike that between "universal" and "particular," admits of degrees (and the idealists, who used it frequently, regularly took the distinction as a relative one).

[1] Read at the meeting of the Western Division of the American Philosophical Association, Madison, Wisconsin, May 2, 1959.

Nevertheless, metaphysicians have traditionally been concerned with the absolutely concrete—that from which absolutely no details have been left out. (They have sometimes been concerned with the absolutely abstract, that from which absolutely all details have been left out—that is to say, Nothing.) A metaphysical theory, indeed, might be defined rather neatly as a theory concerning what is absolutely concrete, i.e., what the world is like with all the details which we men leave out when thinking returned to their proper places. (The abstract, it should be noted, is not "mental" in the sense of being composed of thought; it is what men think *about* when they abstract, and is not to be identified with their thinkings, at least without argument.)

That there has been a confusion in some philosophers' minds between the abstract-concrete distinction and the particular-universal distinction is apparent from the current usage in some quarters of the phrase "the problem of abstract entities." For it is often not clear whether the problem philosophers in those quarters are concerned with is the problem of which entities are abstract and which concrete, or the problem of whether there are repeatable entities. I wish to address myself in what follows to the former problem, while intentionally avoiding the latter. So much to justify the necessity of my opening paragraphs.

Now the theory I am going to propose is that the absolutely concrete is made up of relations, i.e., things which point from something to something else. The immediate response to this will be that relations can't be concrete because they are universals. Russell, for instance, argues in *The Problems of Philosophy* that nominalism can only be considered feasible if one leaves relations out of account; once one remembers them, it is clear that there must be at least one universal—namely, resemblance (*The Problems of Philosophy*, pp. 96–97). But I am not concerned here to argue that the relations I have in mind are particulars, but rather that they are the ultimately concrete entities of nature, or, if you please, that terms referring to them are the "atoms" of a successful metaphysical system.

It may seem to some, however, that the term "relation" is used in philosophy to treat exclusively of abstractions, so that, although relations may or may not be universals, at any rate it is contrary to usage to call them concrete entities. It is for this reason I have decided to speak of the elements of the absolutely concrete as connectors rather than relations.

A relation is ordinarily understood to be an entity which is capable of linking two or more entities. Connectors also are entities which are capable of linking two or more entities. The difference,

then, is that while the term "relation" is ambiguous insofar as it is used to describe entities which may be either abstract or concrete, "connector" will here be used only to describe the absolutely concrete relations.

When I say that relations are capable of linking two or more entities, I am not committing myself to the concrete existence of relations, any more than one who says that being square is a shape some figures have is committing himself to the concrete existence of shapes. It does, of course, commit one to recognizing shape as an abstraction, or at least so I should be willing to admit; but I do not conceive this to be a very exciting admission.

Now it is not necessary, in describing or characterizing a relation, to list the individual entities which are in fact its relata. The same, therefore, is true of connectors. A connector is thus also an entity which is capable of linking suitably ordered entities (pairs, triples, etc.), but which is in addition absolutely concrete. In describing connectors so, I have still not committed myself to their existence. This has to be shown. All I have so far tried to show is that it makes sense to talk about connectors—i.e., that (*a*) it makes sense to talk about relations, conceived as linking entities, and (*b*) it makes sense to talk about concrete entities as distinguished from abstract ones without getting into difficulties about universals.

But it is now necessary, of course, to give examples. I shall argue, then, that in cases of genuine *production*, e.g., where two concrete entities combine to produce a third, the proper analysis of the situation commits us to the concrete existence of connectors—specifically, to at least three: the two entities which combine, and the third, which is not their "combination" but the concrete entity which is produced.

Although I am arguing this view with respect to cases of genuine production, what I am discussing is a part of a more general discussion of causation. Not all cases which have been called instances of causation are genuine cases of production, but some of them are, and I shall begin my treatment by examining two possible and, in my opinion, inadequate theories about the nature of causation in general and production in particular.

A word or two about adequacy. Goodman says, "an adequate system would have to provide a translation for every sentence we care about" (*SA*, p. 21). I tend to take him quite seriously. But it is obvious that philosophers may differ quite widely as to what sentences they care about. One of the sentences I care about is the sentence I shall refer to as (A), "things could have happened otherwise than they did," which I take to assert, unlike its negation,

that there is real freedom and genuine growth in the world. To put it a different way, any metaphysical system I wish to accept will be such that it allows the truth-value of some statements to be contingent upon extra-systematic considerations; a system which makes all statements necessary truths is inadequate for me, since I take the above-quoted sentence seriously, and any such system has no way of translating it. (My reasons for caring about this sentence have to do with the practical contradictions which result from denying it; Kant had similar feelings toward the end of the *Grundlegung.*)

Now there are two theories about the nature of production which seem to me to be inadequate. One of these is the frequently heard version of the empiricist theory which makes the causal relation merely an abstraction. The apparent necessity of the statement "every event has a cause," on this view, stems from our habit of abstracting regularities from the flux of phenomena and then inserting these regularities into our account of the world. There is no necessity in the absolutely concrete, on this view; everything there is accidental, and the appearance of necessity is due to our habits.

Ontologically, this view takes the absolutely concrete (if it takes it at all) to be constituted of bare particulars, uncharacterized in themselves but clothed by our minds with attributes and governed through our abstractings by laws of regular occurrence. Whatever the attractions of this account, however, it will not do as the ontology of my metaphysical system, since it leads eventually to a "closed" universe, i.e., one in which (A) has no use in describing concrete reality. Since the only entities recognized in this account are bare, i.e., uncharacterized particulars, and the characterization of them always involves abstraction, it follows that the characterization of these entities as "happening otherwise than they did" involves an abstraction, that is to say, does not apply to the absolutely concrete. Therefore (A) has no use, unless it is used merely to mean that I could have abstracted differently than I did, and this rather unexciting statement is trivially true, since, as far as abstractions are concerned, to have made them gives them all the being of which they are capable.

The other inadequate view of production is the extreme rationalist position which takes causation as a relationship of logical necessity, such that to know an individual is to know its specific causal relations—the result being that any sentence of the form "*a* causes *b*" is necessarily, because logically, true. It is clear that (A) has no use on this view either, since every specific application of it yields a logical contradiction, in that if *a* had not caused

b it would not be *a* (as we sometimes put it). Thus, this view also leads to a closed universe.

Now I take it as one of the criteria of a successful metaphysical system that it should describe an open universe, i.e., one in whose accurate description (A) or some equivalent statement appears or can be proved. The two inadequate views just examined represent two extremes between which my view will try to steer. On the one hand, the extreme empiricist adopts an ontology whose elements are not characterized at all; on the other hand, the extreme rationalist adopts an ontology whose elements are completely characterized. Unless I am mistaken, then, my view must involve an ontology whose elements are characterized, but incompletely. Again, on the one hand, the radical empiricist's ontology consists of completely unrelated individuals; the radical rationalist's contains completely interrelated individuals. My ontology must therefore contain partly related individuals. Therefore, on my view some relations must be elements of the absolutely concrete, and other relations must be abstract. But this is to say, using the terminology now at hand, that some relations must be connectors, and others not. The problem then becomes, Which relations are the connectors?

The negative arguments against extreme empiricist and rationalist ontologies may serve to give our inquiry plausibility. Now let us try to frame a positive account of an ontology which avoids these extremes.

Let us consider a concrete entity *a* and another concrete entity *b* which are found in the world in such juxtaposition that we find it natural to say that *a* produces *b*. On my view, the metaphysical account of this is that *a* and *b* have come into connection with a third entity, *c*, which connects them. If *a* and *b* are concrete entities, and if we are correct in identifying this sequence as one of genuine production, then it follows on my view that *c* is also a concrete entity.

First off, one might suppose that I hold some sort of view which perversely takes concretion as an additive property, so that if any two things, *x* and *y*, are concrete, then their logical sum, $x + y$, must also be concrete. I hold no such view. The entrance of the concrete entities *a* and *b* into a complex unity, *c*, is not to be conceived on the analogy of two terms entering into their sum. *c* is not the sum of *a* and *b*. Nor, although it is their product, is it their logical product. *c* is a concrete entity, and both the logical sum and product of *a* and *b* are abstract ones.

The next move by the opposition is well-known. I am holding that the relation between two entities is to be explained by appealing

to a third one which relates the first two. But, asks Bradley, what relates each of the first two to the third one? Or am I perhaps willing to countenance an infinite regress? No indeed; that type of infinity makes (A) useless once again for still different reasons—specifically, because nothing could ever happen.

But how, then, shall I avoid the regress? The suggestion is alarmingly simple: it must be that *a* and *b* connect *themselves* to *c*, or, to put it the way I should prefer, that *a* and *b*, as well as *c*, are connectors.

Why does it sound so odd to say that something connects itself to something else? Perhaps it has the ring of a logical paradox; *a* is both connector and connected at the same time and in respect to the same entity, *b*. But I think that ring is spurious. There is no logical contradiction in an entity's being its own connector, as there is, perhaps, in a statement's asserting its own falsity, or in a person's being his own father. Perhaps a part of the ring comes also from the confusion discussed at the outset; it is noticed, in criticizing Platonic views, that in the case of a Form participated in by a particular the relata are of different types, the one universal and the other particular, and it certainly would be a contradiction if one held (*a*) that all relations are universals, (*b*) that a relatum may be its own relation, and (*c*) that some of those relata are particulars, for then those relata would be both universal and particular. But I am not committed to the view that all connectors are universals, nor to its contrary; in fact, I have not advanced any view on the problem of repeatable entities at all.

Actually, when one thinks of it, there are plenty of instances of self-connecting objects in the world of everyday experience. Glue, for example, does not need something else to stick itself to paper, nor does the attraction exerted by a magnet (its field) need something else to connect it to the iron filings. There is nothing hard in thinking of concrete entities as self-connecting; we encounter analogies all the time. Not that I should care to be understood as proposing these as completely satisfactory analogies to the connectors of which I speak; most analogies, being abstractions, will break down if pushed far enough. I merely cite these things to show that the notion of a self-linking connector, far from being self-contradictory or paradoxical, is perfectly well understood and accepted by practically everyone except certain philosophers.

Very well, let us take it as shown that the notion of a self-linking connector is thinkable. Now how is it that a view which takes such connectors as concrete allows what I have called an "open" universe? We saw that to avoid the extremes of empiricism or rationalism our ontology must contain as elements incompletely

characterized entities. My argument must show that self-linking connectors can be understood as incompletely characterized.

An incompletely characterized entity is an entity whose exhaustive characterization involves determinability without involving any specific determinate "lying under" that determinable. Another way of putting this would be to say that concrete entities, being incompletely specified, are of such a nature that some of their specific relations to other concrete entities are not included in their descriptions—i.e., that they have an intrinsic openness about them.

In a case of genuine production, two concrete entities, named *a* and *b*, link themselves to a third concrete entity, *c*, in such a way that we would say that *a* is the cause, *b* the effect, and *c* the productive relation between them. This productive relation might also be described as the concrete situation in which is embedded the abstract *fact* that *b* is the effect of *a*. In other words, *c* is one of the concrete entities from which we abstract our general notion of production. Now, on my view, *a*, *b*, and *c* are all concrete connectors, which is to say they are exhaustively describable by specifying the other sorts of concrete entities (connectors) with which they can connect themselves. Notice I did not say that their description includes the specification of the concrete entities themselves with which they can come into connection, but rather with the *sorts*. Concrete entities are intrinsically incomplete in precisely the sense that their exhaustive characterization involves the use of determinable words rather than determinate ones only.

It is common to retort to such a suggestion that the word "determinable" is significant either (*a*) through its reference to the determinates lying under it, or (*b*) through its reference to a repeatable entity, (*a*) and (*b*) being respectively the views called nominalism and realism. At last, the reader may sigh, I am being forced to face the problem of repeatable entities. Are determinables themselves, then, concrete? The view has its difficulties.

But what has happened in setting up this supposed problem? The critic has made the common mistake of assuming that, because I find myself compelled to use a certain word in ordinary language, I therefore must admit as legitimate the question, "To what concrete thing does that word refer?" If we have learned anything from recent developments in analytic philosophy, it is that this question rests on a mistake. However, I would not want to be understood as saying that I owe no explanation at all of determinability—merely that I owe no argument for or against the existence of concrete determinables simply because I find it expedient to use the language of determinability.

The reason why determinable words are appropriate in char-

acterizing concrete entities is not because there are universals, nor is it the case that in using such words I commit myself to any specific determinate entities falling under the determinable. There might be none. That there *is* intrinsic determinability in the world, however, is precisely what I am arguing, although it doesn't follow from that that I hold that there is an x such that x is either determinability or a determinable.

One may think of a concrete entity, I submit, as a complex of vectors each calling for certain sorts of concrete entities to combine with. These vectors are both negatively and positively determinable, which is to say, they absolutely cannot combine with certain sorts of entities whereas they can or cannot combine with certain other sorts of concrete entities. Expanding this account some more, let me put it this way. A given specific concrete entity, or self-linking connector, consists of vectors which require (but not logically) certain sorts of more complex concrete entities to fill themselves out. Thus *a*, our cause, includes a vector which requires, not *c*, the concrete situation, but rather *a C*, some entity of the sort which is involved when, as we say, things of type A produce things of type B. It is clear that we need different symbols, one to stand for the specific *c* which grows out of *a*'s producing *b*, and one to stand for the *sort C* which A's vector requires. Think of small letters as standing for specific terms and capital letters for sorts. On my view, the analysis of "the 8-ball produced the motion of the 9-ball" is this: given that "*a*" is the name of one of the 8-ball's vectors, and that "*b*" is the name of one of the 9-ball's vectors, that "*c*" is the name of a concrete productive connector, and that "*C*" is the general term for concrete productive connectors, then *a* requires *C*, and *b* requires *C*, and *c* filled out both *a* and *b*.

This account of causality avoids the extreme empiricist view, for it holds that, although *C* is not a concrete entity, *c* is, so that "*a* requires *C*" is a necessary truth, although "*c* filled out *a*" is not. It avoids the extreme rationalist view, for it holds that, although "*a* requires *C*" is a necessary proposition, "*c* filled out *a*" is not.

Although "requires *C*" is linguistically a part of the necessary truth "*a* requires *C*," it does not name or describe a concrete entity in the language rigged up above. The same, then, can be said for "*C*." Either there is no concrete entity *C*, or, if there is, it will have to be shown on different grounds. I believe I have successfully avoided the repeatability issue in this analysis, at least.

The upshot of the discussion, as I see it, is no more than this: some of the traditional bogeys that haunt the incipient realist who believes in freedom can be avoided. This is not to say that there

are no difficulties with the ontology presented here; there clearly are (e.g., how are we to understand the as yet unanalyzed expression "*c* filled out *a*," which we needed in our description?). But it may serve to relocate some of the problems, to remove the onus of traditional difficulties such as the considerations which led Bradley to reject a universe involving concrete relations, for instance, as well as the dilemma of nominalism and realism supposedly posed by the "problem of universals" as it is often formulated.

KARL H. POTTER

UNIVERSITY OF MINNESOTA

ETHICS AND *ETHICS AND THE MORAL LIFE*[1]

IT is sometimes difficult to remember just why people write books on ethics, for they very often fail to make this clear. Mr. Mayo says that the question which "stands behind" all the questions of moral philosophy is this: "How is it that a man's recognition of a moral principle can make a difference to his conduct?" But one's first reaction to this is surely a certain curiosity. Why is this a problem? Would a man be said to have adopted, recognized a certain moral principle unless it *did* make a difference to his conduct? And in fact, many pages later, Mr. Mayo says (p. 114): "When I say that a certain course of action X is wrong, . . . the very least that I am doing . . . is committing myself to a future policy with regard to X. I am, for example, disclosing my readiness to avoid X myself, to resist any inclination towards X, to condemn it on other occasions, to reproach . . . to deter . . . to deplore . . . to try . . . and so on." So presumably if none of these things show themselves in my conduct I haven't really adopted, recognized, the moral principle I enunciated.

Why a book on ethics at all? What *are* the problems in ethics? A philosopher said to me recently: "What I want to examine is the problem whether or not the influencing of decisions or choices is the fundamental role of moral evaluation." This is another (perhaps currently even more popular) way of putting Mr. Mayo's question—i.e., how do moral judgments influence conduct (whether my own or that of another)? Is it perhaps their special distinctive function to do this? But suppose we were to ask who cares how or whether it is. What difference does it make how or whether it is? I shall suggest that these are not trivial questions.

[1] By Bernard Mayo. London: Macmillan (New York: St. Martin's Press), 1958. viii, 238 p.

This page intentionally left blank

DEPENDENT ARISING AND THE EMPTINESS OF EMPTINESS: WHY DID NĀGĀRJUNA START WITH CAUSATION?

Jay L. Garfield

Professor of Philosophy in the School of Communications and Cognitive Science at Hampshire College

1. Introduction

Nāgārjuna, who lived in South India in approximately the first century C.E., is undoubtedly the most important, influential, and widely studied Mahāyāna Buddhist philosopher. He is the founder of the Mādhyamika, or Middle Path, schools of Mahāyāna Buddhism. His considerable corpus includes texts addressed to lay audiences, letters of advice to kings, and the set of penetrating metaphysical and epistemological treatises that represent the foundation of the highly skeptical and dialectical analytic philosophical school known as Mādhyamika. Most important of these is his largest and best-known text, the *Mūlamādhyamikakārikā*—in English, *Fundamental Stanzas on the Middle Way.* This text in turn inspires a huge commentarial literature in Sanskrit, Tibetan, Chinese, Korean, and Japanese. Divergences in interpretation of the *Mūlamādhyamikakārikā* often determine the splits between major philosophical schools. So, for instance, the distinction between two of the three major Mahāyāna philosophical schools, Svātantrika-Mādhyamika and Prāsaṅgika-Mādhyamika, reflect, *inter alia,* distinct readings of this text, itself taken as fundamental by scholars within each of these schools.

The treatise itself is composed in very terse, often cryptic verses, with much of the explicit argument suppressed, generating significant interpretative challenges. But the uniformity of the philosophical methodology and the clarity of the central philosophical vision expressed in the text together provide a considerable fulcrum for exegesis. The central topic of the text is emptiness—the Buddhist technical term for the lack of independent existence, inherent existence, or essence in things. Nāgārjuna relentlessly analyzes phenomena or processes that appear to exist independently and argues that they cannot so exist, and yet, though lacking the inherent existence imputed to them either by naive common sense or by sophisticated, realistic philosophical theory, these phenomena are not nonexistent—they are, he argues, conventionally real.

This dual thesis of the conventional reality of phenomena together with their lack of inherent existence depends upon the complex doctrine of the two truths or two realities—a conventional or nominal truth and an ultimate truth—and upon a subtle and surprising doctrine regarding their relation. It is, in fact, this sophisticated development of the doctrine of the two truths as a vehicle for understanding Buddhist metaphysics and epistemology that is Nāgārjuna's greatest philosophical contribution. If the analysis in terms of emptiness is the substantial heart of *Mūlamādhyamikakārikā,* the method of *reductio ad absurdum* is the methodolog-

ical core. Nāgārjuna, like Western skeptics, systematically eschews the defense of positive metaphysical doctrines regarding the nature of things, demonstrating rather that any such positive thesis is incoherent, and that in the end our conventions and our conceptual framework can never be justified by demonstrating their correspondence to an independent reality. Rather, he suggests, what counts as real depends precisely upon our conventions.[1]

For Nāgārjuna and his followers, this point is connected deeply and directly with the emptiness of phenomena. That is, for instance, when a Madhyamika philosopher says of a table that it is empty, that assertion by itself is incomplete. It invites the question, "empty of what?" And the answer is: "empty of inherent existence, or self-nature, or, in more Western terms, *essence*." Now, to say that the table is empty is hence simply to say that it lacks essence and, importantly, *not* to say that it is completely nonexistent. To say that it lacks essence, the Madhyamika philosopher will explain, is to say, as the Tibetans like to put it, that it does not exist "from its own side"—that its existence *as the object that it is, as a table*, depends not only upon *it* or on any purely nonrelational characteristics, but upon us as well. That is, if this kind of furniture had not evolved in our culture, what appears to us to be an obviously unitary object might instead be correctly described as five objects: four quite useful sticks absurdly surmounted by a pointless slab of stick-wood waiting to be carved. It is also to say that the table depends for its existence on its parts, on its causes, on its material, and so forth. Apart from these, there is no table. The table, we might say, is a purely arbitrary slice of space-time chosen by us as the referent of a single name, and not an entity demanding, on its own, recognition and a philosophical analysis to reveal its essence. That independent character is precisely what it lacks, on this view.

And this analysis in terms of emptiness—an analysis refusing to characterize the nature of any thing, precisely because it denies that we can make sense of the idea of a thing's nature—proceeding by the relentless refutation of any attempt to provide such a positive analysis, is applied by Nāgārjuna to all phenomena, including, most radically, emptiness itself. For if Nāgārjuna merely argued that all phenomena are empty, one might justly indict him for in fact merely replacing one analysis of things with another; that is, with arguing that emptiness is in fact the essence of all things. But Nāgārjuna, as we shall see, argues that emptiness itself is empty. It is not a self-existent void standing behind the veil of illusion represented by conventional reality, but merely an aspect of conventional reality. And this, as we shall see, is what provides the key to understanding the deep unity between the two truths.

While Nāgārjuna is a powerfully original thinker, he is clearly and self-consciously operating squarely within the framework of Buddhist

philosophy. Therefore, Nāgārjuna accepts, and takes it as incumbent upon him, to provide an account of the Four Noble Truths. Moreover, he takes it as a fundamental philosophical task to provide an understanding of what Buddhist philosophy refers to as *pratītyasammutpāda*—dependent co-origination. This term denotes the nexus between phenomena in virtue of which events depend on other events, composites depend upon their parts, and so forth. Just how this dependency is spelled out, and just what is its status is a matter of considerable debate within Buddhist philosophy, just as the nature of causation and explanation is a matter of great dispute within Western philosophy. Nāgārjuna is very much concerned to stake out a radical and revealing position in this debate. I will argue that this position provides the key to understanding his entire text.

The *Mūlamādhyamikakārikā* is divided into twenty-seven chapters. The first chapter addresses dependent origination. While many Western commentators assert that this chapter opens the text simply because it addresses a "fundamental doctrine of Buddhism" (Kalupahana 1986), I will argue that Nāgārjuna begins with causation for deeper, more systematic reasons. In chapters 2 through 23, Nāgārjuna addresses a wide range of phenomena, including external perceptibles, psychological processes, relations, and putative substances and attributes, arguing that all are empty. In the final four chapters, Nāgārjuna replies to objections and generalizes the particular analyses into a broad theory concerning the nature of emptiness itself and the relation between the two truths, emptiness and dependent arising itself. It is generally, and in my view correctly, acknowledged that chapter 24, the examination of the Four Noble Truths, is the central chapter of the text and the climax of the argument. One verse of this chapter, verse 18, has received so much attention that interpretations of it alone represent the foundations of major Buddhist schools in East Asia:

> Whatever is dependently co-arisen
> That is explained to be emptiness.
> That, being a dependent designation
> Is itself the middle way.

Here Nāgārjuna asserts the fundamental identity of (1) emptiness, or the ultimate truth, (2) the dependently originated—that is, all phenomena—and (3) verbal convention. Moreover, he asserts that understanding this relation is itself the middle-way philosophical view he articulates in the *Mūlamādhyamikakārikā*. This verse and the discussion in the chapters that follow provide the fulcrum for Candrakīrti's more explicit characterization of the emptiness of emptiness as an interpretation of Nāgārjuna's philosophical system—the interpretation that is definitive of the Prāsaṅgika-Mādhyamika school. In what follows I will provide an interpretation of this central verse and its context that harmonizes with Candrakīrti's and

argue that, in fact, this doctrine is already to be found in the opening chapter of the text—the examination of conditions. Reading the text in this way, I will argue, locates the doctrine of the emptiness of emptiness not only as a dramatic philosophical conclusion to be drawn at the end of twenty-four chapters of argument, but as the perspective implicit in the argument from the very beginning, and only rendered explicit in chapter 24. Reading the text in this way, I will suggest, also shows us exactly how 24:18 is to be understood, and just why a proper understanding of causality is so central to Buddhist philosophy.

I will begin by offering a philosophical reading of chapter 1. I will argue that Nāgārjuna distinguishes two possible views of dependent origination or the causal process—one according to which causes bring about their effects in virtue of causal powers and one according to which causal relations simply amount to explanatorily useful regularities—and defends the latter. This, I will argue, when suitably fleshed out, amounts to Nāgārjuna's doctrine of the emptiness of causation. I will then turn immediately to chapter 24, focusing on the link between emptiness, dependent origination, and convention, and developing the theory of the emptiness of emptiness. With this in hand, we will return to chapter 1, showing how this doctrine is anticipated in the initial discussion of causation. Finally, I will show quickly how this way of reading the texts changes the way we would read subsequent chapters, and I will make a few general remarks about the moral of this textual exercise for an understanding of the centrality of causation to metaphysics and for an understanding of the remarkably pragmatic outlook of Mādhyamika philosophy.

2. Chapter 1—Examination of Conditions

Central to this first chapter is the distinction between causes and conditions (Skt *hetu* and *pratyaya* [Tib *rGyu* and *rKyen*]. This distinction is variously drawn and is controversial,[2] and is arguably differently understood in Sanskrit and Tibetan. The way I will understand it here, I argue, makes good, coherent sense not only of this chapter, but of the *Mūlamādhyamikakārikā* as a whole. Briefly, we will understand this distinction as follows: When Nāgārjuna uses the word "cause" (*hetu* [*rGyu*]), he has in mind an event or state that has in it a *power* (*kriyā* [*Bya Ba*]) to bring about its effect, and has that power as part of its essence or nature (*svabhāva* [*Rang bZhin*]). When he uses the term "condition," on the other hand (*pratyaya* [*rKyen*]), he has in mind an event, state, or process that can be appealed to in explaining another event, state, or process, without any metaphysical commitment to any occult connection between explanandum and explanans. In chapter 1, Nāgārjuna, we shall see, argues against the existence of causes and for the existence of a variety of kinds of conditions.[3]

The argument against causation is tightly intertwined with the positive account of dependent arising and of the nature of the relation between conditions and the conditioned. Nāgārjuna begins by stating the conclusion (1: 1): neither are entities self-caused nor do they come to be through the power of other entities. That is, there is no causation, when causation is thought of as involving causal activity.[4] Nonetheless, he notes (1: 2), there are conditions—in fact four distinct kinds—that can be appealed to in the explanation and prediction of phenomena. An example might be useful to illustrate the difference between the four kinds of condition, and the picture Nāgārjuna will paint of explanation. Suppose that you ask, "Why are the lights on?" I might reply as follows: (1) Because I flicked the switch. I have appealed to an efficient condition. Or (2) because the wires are in good working order, the bulbs haven't burned out, and the electricity is flowing. These are supporting conditions. Or (3) the light is the emission of photons each of which is emitted in response to the bombardment of an atom by an electron, and so forth. I have appealed to a chain of immediate conditions. Or (4) so that we can see. This is the dominant condition. Any of these would be a perfectly good answer to the "Why?" question. But note that none of them makes reference to any causal powers or necessitation.

The next three verses are crucial. Nāgārjuna first notes (1: 3) that in examining a phenomenon and its relations to its conditions, we do not find that phenomenon somehow contained potentially in those conditions. Now, on the reading of this chapter, I will suggest, we can see conditions simply as useful explanans. Using this language, we can see Nāgārjuna as urging that even distinguishing clearly between explanans and explanandum as distinct entities, with the former containing potentially what the latter has actually, is problematic. What we are typically confronted with in nature is a vast network of interdependent and continuous processes, and carving out particular phenomena for explanation or for use in explanations depends more on our explanatory interests and language than on joints nature presents to us. Through addressing the question of the potential existence of an event in its conditions, Nāgārjuna hints at this concealed relation between praxis and reality.

Next, Nāgārjuna notes (1: 4) that in exploiting an event or entity as a condition in explanation, we do not thereby ascribe it any causal power. Our desire for light does not exert some occult force on the lights. Nor is there anything to be found in the flicking of the switch other than the plastic, metal, movement, and connections visible to the naked eye. Occult causal powers are singularly absent. On the other hand, Nāgārjuna points out in the same breath that this does not mean that conditions are explanatorily impotent. In a perfectly *ordinary* sense—not that which the metaphysicians of causation have in mind—our desire is active in the

production of light. But not in the sense that it contains light potentially, or some special causal power that connects our minds to the bulbs.[5]

What is it, then, about some sets of event pairs, but not others, that make them dependently related, if not some causal link present in some cases but not in others? Nāgārjuna replies (1: 5) that it is the regularities that count. Flickings give rise to illuminations. So they are conditions of them. If they didn't, they wouldn't be. Period. Explanation relies on regularities. Regularities are explained by reference to further regularities. Adding active forces or potentials adds nothing of explanatory utility to the picture.[6]

In reading the next few verses we must be hermeneutically cautious, and pay careful attention to Nāgārjuna's use of the term "existent" (*satah* [*Yod pa*]) and its negative contrastive "nonexistent" (*asatah* [*Med pa*]). For Nāgārjuna is worried here about *inherent* existence and *inherent* nonexistence, as opposed to *conventional* existence or nonexistence. Though this will become clearer as we go along, keep in mind for the present that for a thing to exist inherently is for it to exist in virtue of possessing an essence; for it to exist independently of other entities, and independently of convention. For a thing to be inherently nonexistent is for it to not exist in any sense at all—not even conventionally or dependently.

With this in mind, we can see how Nāgārjuna defends dependent arising while rejecting causation. He notes (1: 6) that if entities are conceived as inherently existent, they exist independently, and hence need no conditions for their production. Indeed, they could not be produced if they exist in this way. On the other hand, if things exist in no way whatsoever, it follows trivially that they have no conditions. This verse and the several that follow (1: 6–10) make this point with regard to each of the four kinds of conditions.

What is important about this strand of the argument? Nāgārjuna is drawing attention to the connection between a causal-power view of causation and an essentialist view of phenomena on the one hand, and between a condition view of dependent arising and a conventional view of phenomena on the other. Here is the point: if one views phenomena as having and as emerging from casual powers, one views them as having essences and as being connected to the essences of other phenomena. This, Nāgārjuna suggests, is ultimately incoherent, since it forces one at the same time to assert the *inherent existence* of these things, in virtue of their essential identity, and to assert their *dependence* and *productive* character, in virtue of their causal history and power. But such dependence and relational character, he suggests, is incompatible with their inherent existence. If, on the other hand, one regards things as dependent merely on conditions, one regards them as merely conventionally existent. And to regard something as merely conventionally existent is to regard it as without essence and without power. And this is to regard it

as existing dependently. This provides a coherent, mundane understanding of phenomena as an alternative to the metaphysics of reification that Nāgārjuna criticizes.

Verse 10 is central in this discussion.

> If things did not exist
> Without essence,
> The phrase, "When this exists so this will be,"
> Would not be acceptable.

Nāgārjuna is replying here to the causal realist's inference from the reality of causal powers to their embodiment in real entities whose essences include those powers. He turns the tables on the realist, arguing that it is precisely because there is no such reality to things—and hence no entities to serve as the bearers of the causal powers the realist wants to posit—that the Buddhist formula expressing the truth of dependent arising[7] can be asserted. It could *not* be asserted if in fact there were real entities. For if they were real in the sense important for the realist, they would be independent. So if the formula were interpreted in this context as pointing to any causal power, it would be false. It can only be interpreted, it would follow, as a formula expressing the regularity of nature.

In the next three verses (1: 11–13) Nāgārjuna anticipates and answers the causal realist's reply. First, the realist argues that the conclusion Nāgārjuna draws from the unreality of causal power—the nonexistence of things (where "existence" is read "inherent existence")—entails the falsity of the claim that things dependently arise (1: 11). For if there are no things, surely nothing arises. This charge has a double edge: if the argument is successful it shows not only that Nāgārjuna's own position is vacuous, but also that it contradicts one of the most fundamental tenets of Buddhist philosophy: that all phenomena are dependently arisen. Moreover, the opponent charges (1: 11), on Nāgārjuna's view that the explanandum is not to be found potentially in the explanans, there is no explanation of how the former is to be understood as depending upon the latter. As Nāgārjuna will emphasize, however (1: 14), the very structure of this charge contains the seeds of its reply. The very emptiness of the effect, an effect presupposed by the opponent to be nonempty, in fact follows from the emptiness of the conditions and of the relationship between conditions and effect. Hence Nāgārjuna can reply to the opponents' attempted refutation by embracing the conclusion of his *reductio* together with the premises it supposedly refutes.

How, the opponent asks, are we to distinguish coincidental sequence from causal consequence? And why (1: 12) don't things simply arise randomly from events that are nonconditions, since no special connection is posited to link consequents to their proper causal antecedents? Finally, the opponent asks (1: 13), since the phenomena we observe clearly have

natures, how could it be, as Nāgārjuna argues, that they proceed by means of a process with no essence, from conditions with no essence? Whence do the natures of actual existents arise? Nāgārjuna again replies to this last charge by pointing out that since on his view the effects indeed have no essence, the opponent's presupposition is ill-founded. This move also indicates a reply to the problem posed in (1: 12); that problem is grounded in the mistaken view that a phenomenon's lack of inherent existence entails that it, being nonexistent, could come into existence from nowhere. But "from nowhere," for the opponent, means from something lacking inherent existence. And indeed, for Nāgārjuna, this is exactly the case: effects lacking inherent existence depend precisely upon conditions which themselves lack inherent existence.

Nāgārjuna's summary of the import of this set of replies (1: 14) is terse and cryptic. But unpacking it with the aid of what has gone before provides an important key to understanding the doctrine of the emptiness of causation that is the burden of this chapter. First, Nāgārjuna points out, the opponent begs the question in asserting the genuine existence of the effects in question. They, like their conditions, and like the process of dependent origination itself, are nonexistent from the ultimate point of view. Hence the third charge fails. As a consequence, in the sense in which the opponent supposes that these effects proceed from their conditions—namely that their essence is contained potentially in their causes, which themselves exist inherently—these effects need not be so produced. And so, finally, the effect-containing conditions for which the opponent charges Nāgārjuna with being unable to account are themselves unnecessary. In short, while the reificationist critic charges the Mādhyamika with failing to come up with a causal link sufficiently robust to link ultimately real phenomena, for the Mādhyamika philosopher, the core reason for the absence of such a causal link is the very absence of such phenomena in the first place.

We are now in a position to characterize explicitly the emptiness of causation, and the way this doctrine is identical with the doctrine of dependent origination from conditions adumbrated in this chapter. It is best to offer this characterization using the *via media* formulation most consonant with Nāgārjuna's philosophical school. We will locate the doctrine as a midpoint between two extreme philosophical views. That midpoint is achieved by taking conventions as the foundation of ontology, hence rejecting the very enterprise of a philosophical search for the ontological foundations of convention (Garfield 1990). To say that causation is nonempty or inherently existent is to succumb to the temptation to ground our explanatory practice and discourse in genuine causal powers linking causes to effects. That is the reificationist extreme which Nāgārjuna clearly rejects. To respond to the arguments against the inherent existence of causation by suggesting that there is then no possibility

of appealing to conditions to explain phenomena—that there is no dependent origination at all—is the extreme of nihilism, also clearly rejected by Nāgārjuna. To assert the emptiness of causation is to accept the utility of our causal discourse and explanatory practice, but to resist the temptation to see these as grounded in reference to causal powers or as demanding such grounding. Dependent origination simply is the explicability and coherence of the universe. Its emptiness is the fact that there is no more to it than that.

Now this is certainly philosophically interesting stuff in its own right. But as I suggested at the outset, there is more to it than just an analysis of causation and dependent arising. For, as we shall see, for Nāgārjuna, among the most important means of demonstrating the emptiness of phenomena is to argue that they are dependently arisen. And so the claim that dependent arising itself is empty will turn out to be the claim that the emptiness of phenomena is itself empty—the central and deepest claim of Mādhyamika ontology.

3. Chapter 24—Examination of the Four Noble Truths

While Chapter 24 ostensibly concerns the Four Buddhist Truths and the way they are to be understood from the vantage point of emptiness, it is really about the nature of emptiness itself, and about the relation between emptiness and conventional reality. As such, it is the philosophical heart of the *Mūlamādhyamikakārikā*. The first six verses of the chapter (24: 1–6) present a reply to Nāgārjuna's doctrine of emptiness by an opponent charging the doctrine with nihilism. The next eight verses (24: 7–14) are primarily rhetorical, castigating the opponent for his misunderstanding of Mādhyamika. The important philosophical work begins with 24: 15. From this point Nāgārjuna offers a theory of the relationship between emptiness, dependent origination, and convention, and argues not only that these three can be understood as co-relative, but that if conventional things (or emptiness itself) were *nonempty*, the very nihilism would ensue with which the reificationist opponent charges Mādhyamika. This tactic of arguing not only against each extreme but also that the contradictory extremes are in fact mutually entailing is a dialectical trademark of Nāgārjuna's philosophical method. Because of the length of this chapter, I will not provide a verse-by-verse reading here, but only a general gloss of the argument, with special attention to critical verses.

The opponent opens the chapter by claiming that if the entire phenomenal world were empty nothing would in fact exist, a conclusion absurd on its face and, more importantly, contradictory to fundamental Buddhist tenets such as the Four Noble Truths (24: 1–6) as well as to conventional wisdom. The implicit dilemma with which Nāgārjuna confronts himself is elegant (24: 6). For as we have seen, the distinction

between the two truths, or two vantage points—the ultimate and the conventional—is fundamental to his own method. So when the opponent charges that the assertion of the nonexistence of such things as the Four Noble Truths and of the arising, abiding, and ceasing of entities is contradictory both to conventional wisdom and to the ultimate truth (namely, on one straightforward interpretation, that all phenomena are impermanent, that is, merely arising, abiding momentarily, and ceasing), Nāgārjuna is forced to defend himself on both fronts and to comment on the connection between these standpoints.

Nāgārjuna launches the reply by charging the opponent with foisting the opponent's own understanding of emptiness on Nāgārjuna. Though this is not made as explicit in the text as one might like, it is important to note that the understanding Nāgārjuna has in mind is one that, in the terms of Mādhyamika, reifies emptiness itself. Verse 24:16 provides a clue.

> If the existence of all things
> Is perceived in terms of their essence,
> Then this perception of all things
> Will be without the perception of causes and conditions.

The opponent is seeing actual existence as a discrete entity with an essence. It would follow that for the opponent, the reality of emptiness would entail that emptiness itself is an entity, and at that an inherently existing entity. To see emptiness in this way is to see it as radically different from conventional, phenomenal reality. It is to see the conventional as illusory and emptiness as the reality standing behind it. To adopt this view of emptiness is indeed to deny the reality of the entire phenomenal, conventional world. It is also to ascribe a special, nonconventional, nondependent hyperreality to emptiness itself. Ordinary things would be viewed as nonexistent, emptiness as substantially existent. (It is important and central to the Mādhyamika dialectic to see that these go together—that nihilism about one kind of entity is typically paired with reification of another.) This view is not uncommon in Buddhist philosophy, and Nāgārjuna is clearly aware that it might be suggested by his own position. So Nāgārjuna's reply must begin by distancing himself from this reified view of emptiness itself and hence from the dualism it entails. Only then can he show that to reify emptiness in this way would indeed entail the difficulties his imaginary opponent adumbrates, difficulties not attaching to Nāgārjuna's own view. This brings us to the central verses of this chapter (24:18 and 24:19):

> Whatever is dependently co-arisen
> That is explained to be emptiness.
> That, being a dependent designation
> Is itself the middle way.

> Something that is not dependently arisen,
> Such a thing does not exist.
> Therefore a non-empty thing
> Does not exist.

These verses demand careful scrutiny. In 24:18, Nāgārjuna establishes a critical three-way relation between emptiness, dependent origination, and verbal convention, and asserts that this relation itself is the Middle Way towards which his entire philosophical system is aimed. As we shall see, this is the basis for understanding the emptiness of emptiness itself. First, Nāgārjuna asserts that the dependently arisen is emptiness. Emptiness and the phenomenal world are not two distinct things. They are rather two characterizations of the same thing. To say of something that it is dependently co-arisen is to say that it is empty. To say of something that it is empty is another way of saying that it arises dependently.

Moreover, whatever is dependently co-arisen is verbally established. That is, the identity of any dependently arisen thing depends upon verbal conventions. To say of a thing that it is dependently arisen is to say that its identity as a single entity is nothing more than its being the referent of a word. The thing itself, apart from conventions of individuation, is nothing but an arbitrary slice of an indefinite spatiotemporal and causal manifold. To say of a thing that its identity is a merely verbal fact about it is to say that it is empty. To view emptiness in this way is to see it neither as an entity nor as unreal—it is to see it as conventionally real. Moreover, "emptiness" itself is asserted to be a dependent designation (Skt *prajñaptir-upadaya* [*brTen Nas gDags pa*]). Its referent, emptiness itself, is thereby asserted to be merely dependent and nominal—conventionally existent but ultimately empty. This is, hence, a middle path with regard to emptiness. To view the dependently originated world in this way is to see it neither as nonempty nor as completely nonexistent. It is, viewed in this way, conventionally existent, but empty. We thus have a middle path with regard to dependent origination. To view convention in this way is to view it neither as ontologically insignificant—it determines the character of the phenomenal world—nor as ontologically efficacious—it is empty. Thus we also have a middle way with regard to convention. And finally, given the nice ambiguity in the reference of "that," (De Ni), not only are "dependent arising" and "emptiness" asserted to be dependent designations, and hence merely nominal, but the very relation between them is asserted to be so dependent, and therefore to be empty.[8]

These morals are driven home in 24:19, where Nāgārjuna emphasizes that everything—and this must include emptiness—is dependently arisen. So everything—including emptiness—lacks inherent existence.

So nothing lacks the three coextensive properties of emptiness, dependent-origination, and conventional identity.

With this in hand, Nāgārjuna can reply to the critic. He first points out (24:20–35) that in virtue of the identity of dependent origination and emptiness on the one hand and of ontological independence and intrinsic reality on the other, such phenomena as arising, ceasing, suffering, change, enlightenment, and so on—the very phenomena the opponent charges Nāgārjuna with denying—are possible only if they are empty. The tables are thus turned: it appears that Nāgārjuna, in virtue of arguing for the emptiness of these phenomena, was arguing that in reality they do not exist, precisely because, for the reifier of emptiness, existence and emptiness are opposites. But in fact, because of the identity of emptiness and conventional existence, it is the reifier who, in virtue of denying the emptiness of these phenomena, denies their existence. And it is hence the reifier of emptiness who is impaled on *both* horns of the dilemma s/he has presented to Nāgārjuna: contradicting the ultimate truth, s/he denies that these phenomena are empty; contradicting the conventional, s/he is forced to deny that they even exist! And so Nāgārjuna can conclude (24:36):

> If dependent arising is denied,
> Emptiness itself is rejected.
> This would contradict
> All of the worldly conventions.

To assert the nonemptiness of phenomena and of their interrelations, Nāgārjuna suggests, when emptiness is properly understood, is not only philosophically deeply confused, it is contradictory to common sense. We can make sense of this argument in the following way: common sense neither posits nor requires intrinsic reality in phenomena or a real causal nexus; common sense holds the world to be a network of dependently arisen phenomena. So common sense holds the world to be empty. Again, the standpoint of emptiness is not at odds with the conventional standpoint, only with a particular philosophical understanding of it—that which takes the conventional to be more than merely conventional. What is curious—and, from the Buddhist standpoint, sad—about the human condition, on this view, is the naturalness and seductiveness of that philosophical perspective.[9]

4. The Emptiness of Emptiness

Let us consider now what it is to say that emptiness itself is empty. The claim, even in the context of Buddhist philosophy, does have a somewhat paradoxical air. For emptiness is, in Mahāyāna philosophical thought, the ultimate nature of all phenomena. And the distinction between the merely conventional nature of things and their ultimate nature would seem to mark the distinction between the apparent and the real.

While it is plausible to say that what is merely apparent is empty of reality, it seems nihilistic to say that what is ultimately real is empty of reality, and, as we have seen, the Mādhyamika are quite consciously antinihilistic. But again, when we say that a phenomenon is empty, we say, *inter alia*, that it is impermanent, that it depends upon conditions, and that its identity is dependent upon convention. Do we really want to say of each phenomenon that its emptiness—the fact that it is empty—is itself impermanent, itself dependent on something else, itself dependent upon conventions? It might at least appear that even if all other properties of conventional entities were so, their emptiness would be an eternal, independent, essential fact.

It may be useful to approach the emptiness of emptiness by first asking what it would be to treat emptiness as nonempty. When we say that a phenomenon is empty, we mean that when we try to specify its essence, we come up with nothing. When we look for the substance that underlies the properties, or the bearer of the parts, we find none. When we ask what it is that gives a thing its identity, we stumble not upon ontological facts but upon conventions. For a thing to be nonempty would be for it to have an essence discoverable upon analysis; for it to be a substance independent of its attributes, or a bearer of parts; for its identity to be self-determined by its essence. A nonempty entity can be fully characterized nonrelationally.

For emptiness to be nonempty would be for it to be a substantial entity, an independent existent, a nonconventional phenomenon. On such a view, arguably held by certain Buddhist philosophical schools, emptiness is entirely distinct from any conventional phenomenon. It is, on such a view, the object of correct perception, while conventional phenomena are the objects of delusive perception. While conventional phenomena are dependent upon conventions, conditions, or the ignorance of obstructed minds, emptiness, on such a view, is apparent precisely when one sees through those conventions, dispels that ignorance, and overcomes those obstructions. It has no parts or conditions, and no properties. Though such a position might appear metaphysically extravagant, it is hardly unmotivated. For one thing, it seems that emptiness does have an identifiable essence—namely the lack of inherent existence. So if to be empty is to be empty of essence, emptiness fails on that count to be empty. Moreover, since all phenomena, on the Mādhyamika view, are empty, emptiness would appear to be eternal and independent of any particular conventions, and hence not dependently arisen. The Two Truths, on such an ontological vision, are indeed radically distinct from one another.

But this position is, from Nāgārjuna's perspective, untenable. The best way to see that is as follows. Suppose that we take a conventional entity, such as a table. We analyze it to demonstrate its emptiness, finding that

there is no table apart from its parts, that it cannot be distinguished in a principled way from its antecedent and subsequent histories, and so forth. So we conclude that it is empty. But now let us analyze that emptiness—the emptiness of the table—to see what we find. What do we find? Nothing at all but *the table's* lack of inherent existence. The emptiness is dependent upon the table. No conventional table—no emptiness of the table. To see the table as empty, for Nāgārjuna, is not to somehow see "beyond" the illusion of the table to some other, more real entity. It is to see the table *as conventional, as dependent.* But the table that we so see when we see its emptiness is the very same table, seen not as the substantial thing we instinctively posit, but rather as it is. Emptiness is hence not different from conventional reality—it is the fact that conventional reality is conventional. Therefore it must be dependently arisen, since it depends upon the existence of empty phenomena. Hence emptiness itself is empty. This is perhaps the deepest and most radical step in the Madhyamika dialectic, but it is also, as we shall see, the step that saves it from falling into metaphysical extravagance and brings it back to sober, pragmatic skepticism.

Now, this doctrine of the emptiness of emptiness emerges directly from 24: 18.

> Whatever is dependently co-arisen
> That is explained to be emptiness.
> That, being a dependent designation
> Is itself the middle way.

For the emptiness of emptiness, as we have just seen, simply amounts to the identification of emptiness with the property of being dependently arisen, and with the property of having an identity just in virtue of conventional, verbal designation. It is the fact that emptiness is no more than this that makes it empty, just as it is the fact that conventional phenomena in general are no more than conventional, and no more than their parts and status in the causal nexus that makes them empty.[10]

So the doctrine of the emptiness of emptiness can be seen as inextricably linked with Nāgārjuna's distinctive account of the relation between the two truths. For Nāgārjuna, as is also evident in this crucial verse, it is a mistake to distinguish conventional from ultimate reality—the dependently arisen from emptiness—at an ontological level. Emptiness just is the emptiness of conventional phenomena. To perceive conventional phenomena as empty is just to see them as conventional, and as dependently arisen. The difference—such as it is—between the conventional and the ultimate is a difference in the way phenomena are conceived/perceived. The point must be formulated with some delicacy, and cannot be formulated without a hint of the paradoxical about it: conventional phenomena are typically represented as inherently existent. We typically

perceive and conceive of external phenomena, ourselves, causal powers, moral truths, and so forth as independently existing, intrinsically identifiable and substantial. But though this is, in one sense, the conventional character of conventional phenomena—the manner in which they are ordinarily experienced—to see them this way is precisely not to see them *as conventional.* To see that they are merely conventional, in the sense adumbrated above and defended by Nāgārjuna and his followers, is thereby to see them as empty, and this is their ultimate mode of existence. These are the two truths about phenomena: On the one hand they are conventionally existent and the things we ordinarily say about them are in fact true, to the extent that we get it right on the terms of the everyday. Snow is indeed white, and there are indeed tables and chairs in this room. On the other hand, they are ultimately nonexistent. These two truths seem as different as night and day—being and nonbeing. But the import of 24: 18 and the doctrine we have been explicating is that their ultimate nonexistence and their conventional existence are the same thing. Hence the deep identity of the two truths. And this is because emptiness is not other than dependent-arising, and hence because emptiness is empty.

Finally, in order to see why chapter 1 is not only an essential groundwork for this central argument, but in fact anticipates it and brings its conclusion to bear implicitly on the whole remainder of the text, we must note that this entire account depends upon the emptiness of dependent origination itself. To see this, suppose for a moment that one had the view that dependent arising were nonempty (not a crazy view, and not *obviously* incompatible with, and arguably entailed by, certain Buddhist doctrines). Then from the identification of emptiness with dependent arising would follow the nonemptiness of emptiness. Moreover, if conventional phenomena are empty, and dependent arising itself is nonempty and is identified with emptiness, then the two truths are indeed two in every sense. Emptiness-dependent arising is self-existent, while ordinary phenomena are not, and one gets a strongly dualistic, ontological version of an appearance-reality distinction. So the argument for the emptiness of emptiness in chapter 24 and the identity of the Two Truths with which it is bound up depend critically on the argument for the emptiness of dependent origination developed in chapter 1.

5. Simple Emptiness versus the Emptiness of Emptiness

We can now see why real causation, in the fully reified cement-of-the-universe sense, as the instantiation of the relation between explanans and explananda could never do from the Mādhyamika standpoint. For though that would at first glance leave phenomena themselves empty of inherent existence, it would retain a nonempty feature of the phenomenal world, and lose the emptiness of emptiness itself. Moreover, a bit of

reflection should lead us to recognize the deep tension in this metaphysics: if the causal powers of things are ultimately real, it is hard to see how one could maintain the merely conventional status of the things themselves. For they could always be individuated as the bearers of those ultimately real causal powers, and the entire doctrine of the emptiness of phenomena would collapse.

Substituting conditions for causes solves this problem. For, as we have seen, by shifting the account in this way we come to understand the relation between conditions and the conditioned as obtaining in virtue of regularity and explanatory utility. And both of these determinants of the relation are firmly rooted in convention rather than in any extraconventional facts. Regularity is always regularity-under-a-description, and descriptions are, as Nāgārjuna puts it, "verbal designations." Explanatory utility is always relative to human purposes and theoretical frameworks. Dependent origination is thus on this model a thoroughly conventional and hence empty alternative to a reified causal model, which nonetheless permits all of the explanatory moves that a theory committed to causation can make. For every causal link one might posit, an equivalent conditional relation can be posited. But the otiose and ultimately incoherent posit of causal power is dispensed with on Nāgārjuna's formulation.

But if the foregoing interpretation is correct, we can make a more radical interpretative claim regarding the structure of *Mūlamādhyamikakārikā*: the entire doctrine of the emptiness of emptiness and the unity of the Two Truths developed in chapter 24 is already implicit in chapter 1. Recall the structure of the argument so far, as we have traced the complex doctrinal web Nāgārjuna spins: the central thesis of chapter 1 as we have characterized it is that there is no inherently existent causal nexus. The link between conditions and the phenomena dependent upon them is empty. To be empty is, however, to be dependent. Emptiness itself is, therefore, as is explicitly articulated in chapter 24, dependent arising. Hence the emptiness of dependent arising is the emptiness of emptiness. And the emptiness of emptiness, as we have seen, is equivalent to the deep identity between the Two Truths. So the entire central doctrine developed in the climactic twenty-fourth chapter is present in embryo in the first. And this is why Nāgārjuna began with causation.

Now, to be sure, it is not apparent on first reading the opening chapter of the *Mūlamādhyamikakārikā* that this is the import of the argument. The rhetorical structure of the text only makes this clear in retrospect, when enough of the philosophical apparatus is on the table to make the entire framework clear. But once we see this framework, a rereading of the text in light of this understanding of the opening chapter is instructive. For it is one thing to argue for the emptiness of some phenomenon *simpliciter* and quite another to argue for that emptiness

with the emptiness of emptiness in mind. If we read the opening chapter in the first way, we are likely to miss the force of many of the particular analyses in the text the depth of which only emerges in light of the deeper thesis of the emptiness of emptiness. If one argues *simply* that a phenomenon is empty of inherent existence, one leaves open the possibility that this is in contrast to phenomena that *are* inherently existent, and hence that the force of this argument is that the phenomenon in question is not *actually* existent. If, on the other hand, one argues that a phenomenon is empty in the context of the emptiness of emptiness, one is explicitly committed to the view that its emptiness does not entail its nonactuality. Emptiness in this context is not nonexistence. The lack of inherent existence that is asserted is not the lack of a property possessed by some entities but not by others, or a property that an entity could be imagined to have, but rather the lack of an impossible attribute. This reorientation of the argument gives what might appear to be a series of starkly nihilistic analyses a remarkably positive tone.

We have time here to consider briefly one example of the difference that this reading of chapter 1 induces in reading the subsequent text. We will consider the analysis of motion and rest in chapter 2. I will not provide a verse-by-verse commentary on the chapter here. But let us note the following salient features of Nāgārjuna's analysis: the target of the argument is a view of motion according to which motion is an entity, or at least a property with an existence independent of that of moving things, or according to which motion is part of the nature of moving things. These are versions of what it would be to think of motion as nonempty. Nāgārjuna argues that from such a view a number of absurd consequences would follow: things not in motion but which were in motion in the past or which will be in the future would have to undergo substantial change, effectively becoming different things when they changed state from motion to rest or vice versa; a regress would ensue from the need for the entity motion itself to be in motion; motion would occur in the absence of moving things; the moment at which a thing begins or ceases motion would be indescribable. Nāgārjuna concludes that a reification of motion is incoherent. Motion is therefore empty.

So far so good. But then, is motion nonexistent? Is the entire universe static according to Mādhyamika philosophy? If we simply read this chapter in isolation, that conclusion might indeed seem warranted. It would be hard to distinguish emptiness from complete nonexistence. We would be left with an illusory world of change and movement, behind which would lie a static ultimate reality. But such a reading would be problematic. For one thing, it would be absurd on its face. Things move and change. Second, it would contradict the doctrine of dependent origination and change that is the very basis of any Buddhist philosophical system, and which Nāgārjuna has already endorsed in the opening chap-

ter. How, then, are we to read this discussion more positively? Answering this question is hermeneutically critical not only for an understanding of this chapter, but—take my word for it—for a reading of the entire text, which, if not read with care, can appear unrelentingly nihilistic. And on such a nihilistic reading, the appearance/reality distinction that is forced can only coincide with the conventional reality/emptiness distinction, resulting in a denial of reality to the mundane world and a reification of emptiness itself.

The positive account we are after emerges when we recall the emptiness of emptiness and read this second chapter in the context of the reinterpreted first chapter: emptiness itself, as we have seen, according to the analysis of dependent arising, is dependently arisen. It is nothing but the emptiness of conventional phenomena, and is the fact of their being dependent and conventional. If emptiness itself is understood as nonempty, on the other hand, then for a phenomenon to be denominated empty is for it to be completely nonexistent. For then its merely conventional character would stand against the ultimate reality of emptiness itself. We have just seen how this would play out in the case of motion, and a moment's reflection would indicate that any other phenomenon subjected to this analysis would fare about as well. But consider, on the other hand, how we interpret the status of motion in light of the emptiness of its emptiness: the conclusion that motion is empty is then simply the conclusion that it is merely conventional and dependent, like the putatively moving entities themselves. Since there is no implicit contrastive, inherently existent ultimate reality, this conclusion does not lead us to ascribe a "second class" or merely apparent existence to motion or to movers. Their nonexistence—their emptiness—is hence itself non-existent in exactly the sense that they are. Existence—of a sort—is thus recovered exactly in the context of an absence of inherent existence.

But existence of what kind? Herein lies the clue to the positive construction of motion that emerges. The existence that emerges is a conventional and dependent existence. Motion does not exist as an entity on this account, but rather as a relation—as the relation between the positions of a body at distinct times, and hence is dependent upon that body and those positions. Moreover, it emerges as a conventional entity in the following critical sense: only to the extent that we make the decision to identify entities that differ from each other in position over time, but are in other respects quite similar, and which form causal chains of a particular sort, as the same entity can we say that the entity so identified moves. And this is a matter of choice. For we *could* decide to say that entities that differ in any respect are thereby distinct. If we did adopt that convention for individuation, an entity here now and one there then would ipso facto be distinct entities. And so no *single entity*

could adopt different positions (or different properties) at different times, and so motion and change would be nonexistent. It is this dependence of motion on the moved, of the status of things as moved on their motion, and of both on conventions of individuation that, on this account, constitutes their emptiness. But this simply constitutes their conventional existence, and provides an analysis of the means by which they so exist. The emptiness of motion is thus seen to be its existence *as conventional* and *as dependent* and hence as not other than its conventional existence. And this just is the emptiness of emptiness. But in understanding its emptiness in this way, we bring motion, change, and movable and changeable entities back from the brink of extinction.

It is thus that seeing Nāgārjuna's analysis of the emptiness of phenomena in the context of the emptiness of emptiness allows for a nonnihilistic, nondualistic, constructive reading of the Mādhyamika dialectic, but a reading which for all of that is rich in its explication of the structure of reality and of our relation to it. But this reading is only accessible in the chapters analyzing particular phenomena if we already find it in chapter 1. And this, I have argued, is possible once we reread that initial chapter in light of the analysis in chapter 24. The Nāgārjuna who emerges is a subtle figure indeed.

6. The Importance of Causation

The analysis of causation can often look like a highly technical aside in philosophy. It might not seem at first glance to be one of the really "big" questions, like those concerning what entities there are, what the nature of mind is, what the highest good is. By contrast, causation often appears to the outsider or to the beginner like one of those recherché corners of philosophy that one has to work one's way into. But of course even in the history of Western metaphysics and epistemology it has always been central. One has only to think of the role of a theory of causation for Hume, Kant, Schopenhauer, or Wittgenstein to see this. This study of the *Mūlamādhyamikakārikā* shows why: a clear understanding of the nature of the causal relation is the key to understanding the nature of reality itself and of our relation to it. For causation is, as Hume, Kant, and Schopenhauer as well as Nāgārjuna emphasize, at the heart of our individuation of objects, of our ordering of our experience of the world, and of our understanding of our own agency in the world. Without a clear view of causation, we can have no clear view of anything.

Nāgārjuna begins by examining the causal relation for this reason generally. But for Nāgārjuna there is a further, more specific reason, one which has no explicit parallel in the work of other systematic philosophers, though it is, to be sure, hinted at darkly in the work of those just mentioned. For Nāgārjuna, by examining the nature of dependent arising, and by showing the emptiness of causation itself, we understand the

nature of emptiness itself, and thereby push the Mādhyamika dialectic of emptiness to its conclusion. By showing causation to be empty, we show all things to be empty, even emptiness itself. Nāgārjuna begins here because, by beginning with causation, the important conclusions he drives at are ready at hand throughout the examination, even if they are not made explicit until much later.

7. Antimetaphysical Pragmatism in Buddhism

When a Westerner first encounters the *Mūlamādhyamikakārikā* or other Mādhyamika texts, the philosophical approach can appear highly metaphysical and downright weird. The unfamiliar philosophical vocabulary, the highly negative dialectic, and the cryptic verse form are indeed forbidding. Most bizarre of all, however, at first glance, is the doctrine that all phenomena, including self and its objects, are empty. For indeed Nāgārjuna and his followers do argue that the entire everyday world is, from the ultimate standpoint, nonexistent. And that does indeed appear to stand just a bit deeper into philosophical left field than even Berkeley dares to play. But if the interpretation I have been urging is adopted, the real central thrust of Mādhyamika is the demystification of this apparently mystical conclusion. While it might appear that the Mādhyamika argue that nothing really exists except a formless, luminous void, in fact the entire phenomenal world, persons and all, are recovered within that emptiness.

And if what I have said is correct, the principal philosophical move in this demystification of emptiness is the attack on a reified view of causality. Nāgārjuna replaces the view shared by the metaphysician and the person-in-the-street—a view that presents itself as common sense, but is in fact deeply metaphysical—with an apparently paradoxical, thoroughly empty, but in the end actually commonsense view not only of causation, but of the entire phenomenal world.

APPENDIX: TRANSLATION OF CHAPTERS 1, 2, AND 24 OF THE *MŪLAMĀDHYAMIKAKĀRIKĀ* (TRANSLATED FROM THE TIBETAN TEXT)

Chapter 1—Examination of Conditions

1. Neither from itself nor from another
   Nor from both,
   Nor from a non-cause
   Does anything whatever, anywhere arise.

2. There are four conditions: efficient condition;
Percept-object condition; immediate condition;
Dominant condition, just so.
There is no fifth condition.

3. The essence of entities
is not evident in the conditions, and so forth.
If these things are selfless,
There can be no otherness-essence.

4. Power to act does not have conditions,
There is no power to act without conditions.
There are no conditions without power to act.
Nor do any have the power to act.

5. These give rise to those,
So these are called conditions.
As long as those do not come from these,
Why are these not non-conditions?

6. For neither an existent nor a nonexistent thing
Is a condition appropriate.
If a thing is nonexistent, how could it have a condition?
If a thing is already existent, what would a condition do?

7. Neither existents nor
Nonexistents nor existent nonexistents are produced.
In this case, how would there be a "productive cause?"
If it existed, how would it be appropriate?

8. Certainly, an existent mental episode
Has no object.
Since a mental episode is without an object,
How could there be any percept-condition?

9. Since things are not arisen,
It is not acceptable that they cease.
Therefore, an immediate condition is not reasonable.
If something has ceased, how could it be a condition?

10. If things did not exist
Without essence,
The phrase, "When this exists so this will be,"
Would not be acceptable.

11. In the various conditions united,
The effect cannot be found.
Nor in the conditions themselves.
So how could it come from the conditions?

12. However, if a nonexistent effect
Arises from these conditions,
Why does it not arise
From non-conditions?

Jay L. Garfield

13. If the effect is the conditions' essence,
Then the conditions do not have their own essence.
So, how could an effect come
From something that is essenceless?

14. Therefore, conditions have no essence.
If conditions have no essence, there are no effects.
If there are no effects without conditions,
How will conditions be evident?

Chapter 2—Examination of Motion

1. What has been moved is not moving.
What has not been moved is not moving.
Apart from what has been moved and what has not been moved,
Movement cannot be conceived.

2. Where there is flux, there is motion.
Since there is flux in the moving,
And not in the moved or not-moved,
Motion is in that which is moving.

3. If motion is in the mover,
Then how would it be acceptable
When it is not moving,
To have called it a mover?

4. The motion of what moves?
What motion does not move?
Given that that which has passed is gone,
How can motion be in the moved?

5. If motion is in the mover,
There would have to be a twofold motion:
One in virtue of which it is a mover,
And one in virtue of which it moves.

6. If there were a twofold motion,
The subject of that motion would be twofold.
For without a subject of motion,
There cannot be motion.

7. If there is no mover
It would not be correct to say that there is motion.
If there is no motion,
How could a mover exist?

8. Inasmuch as a real mover does not move,
And a nonmover does not move,
Apart from a mover and a nonmover,
What third thing could move?

9. When without motion,
It is unacceptable to call something a mover,
How will it be acceptable
To say that a moving thing moves?

10. For him from whose perspective a mover moves,
There is no motion.
If a real mover were associated with motion,
A mover would need motion.

11. If a mover were to move,
There would be a twofold motion:
One in virtue of which he is a mover,
And one in virtue of which the mover moves.

12. Motion does not begin in what has moved,
Nor does it begin in what has not moved,
Nor does it begin in what is moving.
In what, then, does motion begin?

13. If motion was begun in the past,
When should we say it began?
Not in the nongoing, not in the gone.
How could it be in the nonmoved?

14. Since the beginning of motion
Cannot be conceived,
What gone thing, what going thing,
And what nongoing thing can be conceived?

15. A moving thing is not at rest.
A nonmoving thing is not at rest.
Apart from the moving and the nonmoving,
What third thing is at rest?

16. If without motion
It is not appropriate to posit a mover,
How could it be appropriate to say
That a moving thing is stationary?

17. One does not halt from moving,
Nor from having moved or not having moved.
Motion and coming to rest
And starting to move are similar.

18. That motion is the mover
Itself is not correct.
Nor is it correct that
They are different.

19. It would follow from
The identity of mover and motion
That agent and action
Are identical.

Jay L. Garfield

20. It would follow from
A real distinction between motion and mover
That there could be a mover without motion
And motion without a mover.

21. When neither in identity
Nor in difference,
Can motion and the mover be established as existent,
How can they be established as entities at all?

22. The motion by means of which a mover is manifest
Cannot be the motion by means of which he moves.
He does not exist before that motion,
So what and where is the thing that moves?

23. A mover does not carry out a different motion
From that by means of which he is manifest as a mover.
Moreover, in one mover
A twofold motion is unacceptable.

24. A really existent mover
Does not move in any of the three ways.
A nonexistent mover
Does not move in any of the three ways.

25. Neither an entity nor a nonentity
Moves in any of the three ways.
So movement and motion
And Agent of motion are nonexistent.

Chapter 24—Examination of the Four Noble Truths

1. If all of this is empty,
Not arising, abiding, or ceasing,
Then for you, it follows that
The Four Noble Truths do not exist.

2. If the Four Noble Truths do not exist,
Then knowledge, abandonment,
Meditation, manifestation, and action
Will be completely impossible.

3. If these things do not exist,
The four fruits will not arise.
Then there will not be the enterers into the path.
If not, there will not be the eight [kinds of practitioner].

4. If so the assembly of holy ones
Itself will not exist.
If the Four Noble Truths do not exist,
There will be no true Dharma.

5. If there is no doctrine and assembly
How can there be a Buddha?
If emptiness is conceived in this way
The Three Jewels are contradicted.

6. The attainment of the real fruits
And the Dharma will not exist, and the Dharma itself
And the conventional truth
Will be contradicted.

7. This understanding of yours
Of emptiness and the purpose of emptiness
And of the significance of emptiness is incorrect.
As a consequence you are harmed by it.

8. The Buddha's teaching of the Dharma
Is based on two truths:
A truth of worldly convention
And an ultimate truth.

9. Those who do not understand
The distinction drawn between these two truths
Do not understand
The Buddha's profound truth.

10. Without a foundation in the conventional truth
The significance of the ultimate cannot be taught.
Without understanding the significance of the ultimate,
Liberation is not achieved.

11. By a misperception of emptiness
A person of little intelligence is destroyed.
Like a snake incorrectly seized
Or like a spell incorrectly cast.

12. For that reason—that the Dharma is
Deep and difficult to understand and to learn—
That (the Buddha's) mind despaired of
Being able to teach it.

13. If a fault in understanding should arise
with regard to emptiness, that would not be good.
Your confusion about emptiness, however,
Would not belong to me.

14. For him to whom emptiness is clear,
Everything becomes clear.
For him for whom emptiness is not clear,
Nothing becomes clear.

15. If you foist on us
All of your divergent views
Then you are like a man who has mounted his horse
And has forgotten that very horse.

16. If the existence of all things
Is perceived by you in terms of their essence,
Then this perception of all things
Will be without the perception of causes and conditions.

17. Effects and causes
And agent and action
And conditions and arising and ceasing
And effects will be rendered impossible.

18. Whatever is dependently co-arisen
That is explained to be emptiness.
That, being a dependent designation
Is itself the middle way.

19. Something that is not dependently arisen,
Such a thing does not exist.
Therefore a nonempty thing
Does not exist.

20. If all this were nonempty, as in your view,
There would be no arising and ceasing.
Then the Four Noble Truths
Would become nonexistent.

21. If it is not dependently arisen,
How could suffering come to be?
Suffering has been taught to be impermanent,
And so cannot come from its own essence.

22. If something comes from its own essence,
How could it ever be arisen?
It follows that if one denies emptiness
There can be no arising [of suffering].

23. If suffering had an essence,
Its cessation would not exist.
So if an essence is posited
One denies cessation.

24. If the path had an essence,
Cultivation would not be appropriate.
If this path is indeed cultivated,
It cannot have an essence.

25. If suffering, arising, and
Ceasing are nonexistent,
If through the path suffering ceases,
In what way could one hope to attain it?

26. If through its essence
non-understanding comes to be,
In what way will understanding arise,
Is not essence stable?

27. In this way you should understand
the activities of relinquishing and realizing and
Cultivation and the Four Fruits.
It [essence] is not appropriate.

28. For an essentialist,
Since the fruits through their essence
Are already realized
In what way could it be appropriate to cultivate them?

29. Without the fruits, there are no attainers of the fruits,
Or enterers into that stream,
From this it follows that the eight kinds of persons do not exist.
If these do not exist, there is no spiritual community.

30. From the nonexistence of the Noble Truths
Would follow the nonexistence of the True Doctrine.
If there is no Doctrine and no Community,
How could a Buddha arise?

31. Your enlightened Buddha,
Without relying on anything, would have come to be;
Your Buddha's enlightenment,
Without relying on anything, would have come to be.

32. If by means of your essence
Someone were unenlightened,
Even by practicing towards enlightenment
He could not achieve enlightenment.

33. With neither entities nor nonentities
There can be no action.
What could the nonempty do?
With an essence there is no action.

34. With neither entities nor nonentities
The fruit would arise for you.
So, for you a fruit caused by entities or nonentities
Could not arise.

35. If, for you, a fruit
Were given rise to by either entities or nonentities,
Then from entities or nonentities
How could a nonempty fruit arise?

36. If dependent arising is denied,
Emptiness itself is rejected.
This would contradict
All of the worldly conventions.

37. If emptiness itself is denied,
No action will be appropriate.
Action would not begin,
And without action there would be no agent.

38. If there is essence, all of the flux
Will be unarising, unceasing,
And static. And so, the entire sphere of
Various arisen things would be nonempty.

39. If the empty does not exist,
Then action will be without profit.
The act of ending suffering and
Abandoning misery and defilement will not exist.

40. Whoever sees dependent arising
Also sees Suffering
And Misery and its arising
And the path to its cessation.

NOTES

Thanks are extended to the Venerable Lobzang Norbu Shastri and Janet Gyatso for a very thorough critical reading of and helpful critical comments on an earlier draft of this essay and of the relevant fragments of the translation, and to G. Lee Bowie and Meredith Michaels for sound suggestions regarding that draft. This essay has also benefited from the insightful questions posed by an audience at Mount Holyoke College, and from the sound suggestions of Tom Wartenberg on that occasion. My deepest appreciation goes to the Venerable Geshe Yeshes Thap-Kas for his patient and lucid teaching of this text and discussion of Nāgārjuna's position, and to the Central Institute of Higher Tibetan Studies, to its director the Venerable Professor Samdhong Rinpoche, and to my many colleagues there, including those just mentioned and the Venerable Ngawang Samden and the Venerable Geshe Ngawang Sherab. Thanks also to my research assistant both at the Institute and at Hampshire College, Sri Yeshe Tashi Shastri, and to the Indo-American Fellowship program for grant support while I was working on these ideas.

1 – A fine point, suggested by Janet Gyatso: Though in the end, as we shall see, ultimate reality depends on our conventions in a way, it depends on our conventions in a very different way from that in which conventional reality does. Despite this difference in the structure of the relation between convention and reality in the two cases, however, it remains a distinctive feature of Nāgārjuna's system that it is impossible to speak coherently of reality independent of conventions.

2 – Some argue that there is no real difference between causes and conditions; some that a cause is one kind of condition; some that

efficient causes are causes, and that all other causal factors contributing to an event are conditions. Some like my reading. I have found no unanimity on this interpretative question, either among Western Buddhologists or among Tibetan scholars. The canonical texts are equivocal as well. I do not argue that the distinction I here attribute to Nāgārjuna, which I defend on hermeneutical grounds, is necessarily drawn in the same way throughout the Buddhist philosophical world, or even throughout the Prāsaṅgika-Mādhyamika literature. But it is the one Nāgārjuna draws.

3 – There are two kinds of case to be made for attributing this distinction to Nāgārjuna in this chapter. Most generally, there is the hermeneutical argument that this makes the best philosophical sense of the text. It gets Nāgārjuna drawing a distinction that is clearly suggested by his philosophical outlook and that lines up nicely with the technical terms he deploys. But we can get more textually fine-grained as well: in the first verse, Nāgārjuna explicitly rejects the existence of efficacy, and pointedly uses the word "cause." He denies that there are such things. Nowhere in chapter 1 is there a parallel denial of the existence of conditions. On the contrary, in verse 2 he positively asserts that there are four kinds of them. To be sure, this could be read as a mere partitioning of the class of effects that are described in Buddhist literature. But there are two reasons *not* to read it thus. First, Nāgārjuna does not couch the assertion in one of his "It might be said" locutions. Second, he never takes it back. The positive tone the text takes regarding conditions is continued in verses 4 and 5, where Nāgārjuna asserts that conditions are conceived without efficacy *in contrast with the causes rejected in 1*, and where he endorses a regularist view of conditions. So it seems that Nāgārjuna does use the "cause"/"condition" distinction to mark a distinction between the kind of association he endorses as an analysis of dependent arising and one he rejects.

4 – The Venerable Lobzang Norbu Shastri has pointed out to me that this verse may not in fact be original with Nāgārjuna, but is a quotation from sutra. It appears in the *Kamsika-prajñapāramitāsūtra* as well as in the *Mādhyamika-Śālistambasūtra*. Inasmuch as these are both late texts, their chronological relation to Nāgārjuna's text is not clear.

5 – There is also a nice regress to be developed here that Nāgārjuna does not explicitly note in this chapter, though he does make use of it later in the *Mūlamādhyamikakārikā* (chap. 7): Even if we did posit a causal power mediating between causes and their effects, we would have to explain how it is that a cause event gives rise to or acquires that power, and how the power brings about the effect. We now have two nexuses to explain, and now each one has an

unobservable entity on one end. In Garfield 1990 I explore this problem in more detail and note that it is explored both by Hume and by Wittgenstein in the *Tractatus*.

6 – The Mādhyamika position implies that we should seek to explain regularities by reference to their embeddedness in other regularities, and so on. To ask why there are regularities at all, on such a view, would be to ask an incoherent question: the fact of explanatorily useful regularities in nature is what makes explanation and investigation possible in the first place, and is not something itself that can be explained. After all, there is only one universe, and truly singular phenomena, on such a view, are inexplicable in principle. This may connect deeply to the Buddha's insistence that questions concerning the beginning of the world are unanswerable.

7 – A formula familiar in the sutras of the Pali canon.

8 – Though this is beyond the scope of this essay, this last fact, the emptiness of the relation between the conventional world of dependently arisen phenomena and emptiness itself is of extreme importance at another stage of the Mādhyamika dialectic, and comes to salience in the *Vigrahavyāvartanī* and in Candrakīrti's *Prasannapadā*. For this amounts to the emptiness of the central ontological tenet of Nāgārjuna's system, and is what allows him to claim, despite all appearances, that he is positionless. That is, Nāgārjuna thereby has a ready reply to the following apparent *reductio* argument (reminiscent of classical Greek and subsequent Western challenges to Pyrrhonian skepticism): You say that all things are, from the ultimate standpoint, nonexistent. That must then apply to your own thesis. It, therefore, is really nonexistent, and your words are hence only nominally true. Your own thesis, therefore, denies its own ground and is self-defeating. This objection would be a sound one against a view that in fact asserted its own inherent existence, or grounded its truth on an inherently existing ontological basis. But, Nāgārjuna suggests here, that is not the case for his account. Rather, on his analysis, everything, including this very thesis, has only nominal truth, and nothing is either inherently existent, or true in virtue of designating an inherently existent fact.

9 – This, of course, is the key to the soteriological character of the text: reification is the root of grasping and craving, and hence of all suffering. And it is perfectly natural, despite its incoherence. By understanding emptiness Nāgārjuna intends one to break this habit and extirpate the root of suffering. But if in doing so one falls into the abyss of nihilism, nothing is achieved. For then, action itself is impossible and senseless, and one's realization amounts to nothing. Or

again, if one relinquishes the reification of phenomena but reifies emptiness, that issues in a new grasping and craving—the grasping of emptiness and the craving for nirvāṅa—and a new round of suffering. Only with the simultaneous realization of the emptiness but conventional reality of phenomena and of the emptiness of emptiness, argues Nāgārjuna, can suffering be wholly uprooted.

10 – Paradox may appear to loom at this point. For, one might argue, if emptiness is empty, and if to be empty is to be merely conventional, then the emptiness of any phenomenon is a merely conventional fact. Moreover, to say that entities are merely conventional is merely conventional. Hence it would appear optional, as all conventions are, and it would further seem to be open to say that things are in fact nonconventional, and therefore nonempty. This would be a deep incoherence indeed at the heart of Nāgārjuna's system. But the paradox is merely apparent. The appearance of paradox derives from seeing "conventional" as functioning logically like a negation operator—a subtle version of the nihilistic reading Nāgārjuna is at pains to avoid, with a metalinguistic twist. For then, each iteration of "conventional" would cancel the previous occurrence, and the conventional character of the fact that things are conventional would amount to the claim that really they are not, or at least that they might not be. But in Nāgārjuna's philosophical approach, the sense of the term is more ontological than logical: to say of a phenomenon or of a fact that it is conventional is to characterize its mode of subsistence. It is to say that it is without an independent nature. The fact that a phenomenon is without independent nature is, to be sure, a further phenomenon—a higher-order fact. But that fact, too, is without an independent nature. It, too, is merely conventional. This is another way of putting the strongly nominalistic character of Mādhyamika philosophy. So, a Platonist, for instance, might urge (and the Mādhyamika would agree) that a perceptible phenomenon is ultimately unreal. But the Platonist would assert that its properties are ultimately real. And if some Buddhist-influenced Platonist would note that among the properties of a perceptible phenomenon is its emptiness and its conventional reality, s/he would assert that these, as properties, are ultimately real. This is exactly where Nāgārjuna parts company with all forms of realism. For he gives the properties a nominalistic construal, and asserts that they, including the properties of emptiness and conventionality, are, like all phenomena, merely nominal, merely empty, and merely conventional. And so on for their emptiness and conventionality. The nominalism undercuts the negative interpretation of "conventional" and so renders the regress harmless.

REFERENCES

Garfield, Jay L. 1990. "Epoche and Sūnyatā: Skepticism East and West," *Philosophy East and West* 40:285–307; reprinted in Glazer and Miller, eds., *Words that Ring Clear as Trumpets.* Amherst: Hampshire College Press, 1992.

Kalupahana, David. 1986. *Nāgārjuna: The Philosophy of the Middle Way.* Albany: State University of New York Press.

KARL H. POTTER

Freedom and Determinism from an Indian Perspective[*]

IT IS A COMMON misconception that the commitment of Indians to the so-called "law of karma," transmigration, and the beginninglessness of the self involves them in some sort of fatalism. In a paper delivered at the Fourth East-West Philosophers' Conference in Hawaii in 1964, Kalidas Bhattacharyya has once again emphasized that the texts show that Indian philosophers accept no kind of fatalism. Not only do practically all the Indian systems accept the possibility of the ultimate liberation of each and every self, but they also vehemently affirm the freedom of every individual here and now to choose and act according to his choice.

Nevertheless, the doctrine of karma does involve a fairly strong form of determinism. According to Bhattacharyya,

> . . . the Law of *Karma,* properly understood, is *not* against freedom of action. It is true that, according to this law, my present life is somehow determined by the merits and demerits of what I did in my previous life, that life, again, by the merits and demerits of what I had done in a life still previous, and so on ad infinitum, but this does not mean that everything of a particular life is determined. It has been explicitly stated that only three things in my life are determined. They are (1) the bodily and mental make-up and social position at the time of birth in that particular life (*jāti*), (2) the span of that life (*āyuḥ*), and (3) particular experiences with all the hedonic tones that they have in that life and all that is necessary as objects or direct or indirect causes of those experiences (*bhoga*). Among these direct or indirect causes are included my physiological movements and actions which proceed through attachment and repulsion but not actions at the moral and spiritual levels.

For

> moral actions . . . channelizing . . . normal biological activities are free in the sense that . . . there is no blind submission [to the emotions]. . . . Actions which proceed

[*] This paper was read at the meeting of the American Philosophical Association's Western Division at Minneapolis, Minnesota, May, 1966.

consciously through attachment and repulsion and are consequently accompanied by an I-feeling are free only in the sense that in their case there is always a possibility of not having submitted to them.[1]

In other words, the person, of beginningless origins, matures by moving successively from stages of his being where his movements are completely controlled by passions to a stage where he "channelizes" his emotional responses, and finally to a stage where, through detachment from the passions, he is completely free and in control of himself.

Just how "hard" a determinism Bhattacharyya is specifying is not altogether clear, and it is the aim of this paper to try to clarify it. I shall suggest that we regularly confuse determinism with determinateness, and that what Bhattacharyya means to propound as the Hindu view of freedom and determinism is a position perfectly compatible with determinism, although involving indeterminacy of some events, namely those he says are not "determined" —moral and spiritual actions. Moreover, I shall suggest that the distinction between determinacy and determinism is one which we can utilize to advantage in resolving the issues surrounding the notion of freedom, provided we accept, as the Hindus do, the hypothesis of the beginninglessness of the agent-self. But, as I see it, this is a price we must be prepared to pay whether we accept determinism or not.

Let us begin with the notion of the beginninglessness of the self. This strange-sounding hypothesis is, it has been argued, entailed by the joint acceptance of determinism and the existence of morally responsible agents. The argument was set forth by John Wisdom in 1934[2] and I have attempted[3] to bring it up to date by deriving the conclusion, from recent analysis of "having an ability,"[4] that a responsible agent must be beginningless.

My argument, summarized briefly, is as follows: according to contemporary analyses, "A has the ability to x" is to be rendered as "A is in a condition such that, given opportunity, if he tries to x then he succeeds a certain percentage of the time." In applying this to a particular case, we will want to identify times as t_1, t_2, etc., and it will be convenient if the higher-numbered

[1] Kalidas Bhattacharyya, "The Status of the Individual in Indian Metaphysics," in Charles A. Moore, ed., *The Indian Mind* (Honolulu: East-West Center Press and University of Hawaii Press, 1967), pp. 302-303.

[2] John Wisdom, *Problems of Mind and Matter* (Cambridge, England: Cambridge University Press, 1934), pp. 123-126.

[3] In a paper which was read at the Conference on the Self held at Wooster College in March, 1965, to be published in a volume of proceedings emanating from that conference.

[4] Such an analysis is given by Arnold S. Kaufman in his paper "Ability," *Journal of Philosophy* 60 (1963), 537-551.

times are construed to be earlier than the lower-numbered ones. Then a plausible analysis along the lines suggested of "Smith has the ability to stamp-his-foot-at-t_1" will be "There is an independently specifiable condition C such that, for any time t_n, if at t_{n+1} Smith is in C and, given opportunity, tries at t_{n+1} to stamp his foot, then he succeeds in stamping his foot at t_n a certain percentage of the time; and Smith is in C at t_2.[5] Now, furthermore, if Smith has the ability to stamp his foot at t_1 it must also be true that he has the ability to try to stamp his foot at t_2. Since trying to do something is as much an action as doing something, we can proceed to analyze "Smith has the ability to try at t_2 to stamp-his-foot-at-t_1" in a fashion parallel to that in which we have just analyzed "Smith has the ability to stamp-his-foot-at-t_1"; the analysis of "Smith has the ability to try at t_2 to stamp-his-foot-at-t_1" will be "There is a condition C′ such that, for any time t_n, if at t_{n+1} Smith is in C′ and, given opportunity, tries to try-at-t_n-to-stamp-his-foot-at-t_{n-1}, then he succeeds in trying-at-t_n-to-stamp-his-foot-at-t_{n-1} a certain percentage of the time; and Smith is in C′ at t_3."[6]

If the moves I have made to this point are allowed, it is easy to see that a regress is under way. Furthermore, if we reflect that to attribute responsibility to Smith for stamping his foot requires our likewise attributing to him the ability to stamp his foot, and that in like fashion if he is responsible for trying to stamp his foot he must be able to try to stamp his foot, it will become apparent that if in general the ability to try to x is a prerequisite of our attributing responsibility to someone for x-ing, the agent must be beginningless. For if he is not beginningless, there is some action of his which he is unable to try to perform and yet his ability to perform it is a necessary condition for his being responsible for his performance of x. In short, if the agent is not beginningless he cannot be responsible for any of his actions—at least on the accounts of responsibility and of having abilities which I have adopted in the argument.

No doubt a characteristic response to this will be that either something has gone wrong with the argument or else, if it is indeed valid, it serves as a *reductio ad absurdum* of the presupposed analysis of "having an ability" or of the presumed account of responsibility. Since the validity of the argument is the subject of another paper I shall not argue out objections to it here. I think I can take care of the objections I have found in the literature. But I

---

[5] Kaufman, *ibid.*, suggests that the state C might be thought of as physiological, but this is not a necessary assumption.

[6] Although I grant that the phrase "tries to try" is not used, I think this fact can be explained consonantly with the thesis that "trying to x" denotes an action distinct from x-ing. The argument is contained in the paper referred to in footnote 3.

am not inclined to take the argument as a *reductio* of the analyses of ability or of responsibility, for I am not at all convinced that it is possible to provide any suitable alternative analyses. In any case, what I want the reader to do now is to accept for the moment the possibility that moral agents are beginningless—which is what the Indians hold as a matter of course—and to ask whether, given this assumption, a fresh and promising start can be made on that very vexed question: how to reconcile the agent's freedom with determinism?

Now in the first place, the beginninglessness of agents certainly appears to remove one serious obstacle to squaring freedom with determinism, since it makes it possible to deny that all the causes of the agent's actions are ultimately traceable "outside" him. But this is a rather abstract advantage, for it is not at all clear that the most severe difficulties about freedom and determinism turn on the "spatio-temporal" location of the ultimate causes of actions. What, then, are the more severe difficulties?

Out of the reams of discussion about this point I wish to select for study here the following formulation of the problem: how, within a deterministic context, can we make sense of the phrase "could have done otherwise" in "A x-ed, but he could have done otherwise, so he was free and responsible" and "A x-ed, but he could not have done otherwise, so he is not responsible because he was not free." This seems to be a sufficiently profound way to pose the issues; indeed, philosophers have been in such despair over solving this question that they have regularly turned to an attempt to qualify or modify the notions of freedom and responsibility, or else they have explicitly or implicitly admitted some sort of indeterminism.

At first glance it might seem that the analysis of "having an ability" itself constitutes an analysis of "could have done otherwise." But to say that A could have done otherwise than x is not only to say that he had the ability to x but also to say that he had the ability to do something other than x. Just how to formulate this is something of a puzzle, as we shall see. Nevertheless, the line looks promising. For if A may have had the ability to x without it necessarily being the case also that he could have done otherwise, then we have a way of distinguishing A's actions from his mere behavior (A's actions are those he has the ability to do and, given opportunity, does, as opposed to things which just happen to him) as well as a way of distinguishing necessary conditions for actions for which A is responsible from those which are just his actions (A is only responsible for actions which he has the ability not to perform as well as to perform). As a result A can do actions for which he is not morally responsible as well as those for which he is, and this en-

ables us to adopt our common practices of social and legal discriminations —praising or excusing—free from the suspicion that these practices rest on some ultimate confusion.

The promising line, then, equates "A could have done otherwise" with "A had the ability either to x or to do something other than x." Utilizing the expansion of ability-talk suggested previously, this will come to something like the following:

A could have done otherwise than x at t = There is a condition C such that for any time t_n, if at t_{n+1} A is in C and, given opportunity both to x and to $\bar{x}$ tries to x, then he succeeds in x-ing at t_n a certain percentage of the time; but if at t_{n+1} A is in C and, given opportunity both to x and to $\bar{x}$, tries to $\bar{x}$, then he succeeds in $\bar{x}$-ing at t_n a certain percentage of the time; and at t + 1 A was in condition C and had opportunity both to x and to $\bar{x}$.

In the above, the expression "x" stands for a set of actions of a specifiable sort, and the expression "$\bar{x}$" stands for a set of actions such that an agent cannot perform both x and $\bar{x}$ at the same time. Thus, to take an example, to say that A could have done otherwise than stamp his foot at t is, on the analysis given, to say that there is a condition C such that, for any time t_n, if at t_{n+1} A is in C and, given opportunity to stamp his foot or to do some other relevant action as an alternative, tries to stamp his foot at t_n, then he succeeds in stamping his foot at t_n a certain percentage of the time; but if at t_{n+1} A is in C and, given the same sort of opportunity, tries to do one of the other relevant actions, then he succeeds in that alternative action at t_n a certain percentage of the time; and A was in condition C at t and the opportunity of the sort mentioned was presented to him at t.

It is evident that the ranges covered by pairs of alternative sets of actions x and $\bar{x}$ may vary according to cases. But whatever the scope of $x + \bar{x}$, to say that A could have done otherwise than perform x is in effect to say that A has a complex ability, the ability-to-x-or-$\bar{x}$, and that the circumstances do not preclude either action.

The next question to consider is whether this account violates determinism. Are there necessarily any uncaused events in the story developed here? Suppose that A tried and succeeded in doing x at t_1, although he could have done otherwise; was there a sufficient condition for his doing x? Yes; it consisted in his having the opportunity at t_2 to x or $\bar{x}$, having the ability to x or $\bar{x}$, and trying to x. Each of these three causal factors is itself an event. Does each in turn have causes? Suppose we can develop an extended story on

the basis of which we can answer "yes" in each of the three cases by appealing to prior events, and so on indefinitely. This regress is no longer in itself a defect, for *ex hypothesi* we have accepted beginninglessness.

For example, take A's trying to x at t_2. What is its cause, a sufficient condition for it? It seems to me that "tries" denotes an action; it follows, if that is the case, that *if* A could have done otherwise than to try at t_2 to x there have occurred three causal factors (of the sort specified before) jointly sufficient to produce A's trying at t_2 to x. But it is not necessary that A could have done otherwise than to try at t_2 to x, even though he could have done otherwise than to x at t_1. That is, it is quite possible that A could not have done otherwise than try to x at $t + 1$, though he could have done otherwise than x at t. E.g., he might have been compelled to try to stamp his foot at $t + 1$, but quite able to fail to do so because, say, he failed to keep his ankle firm and so succeeded not in stamping his foot but only in turning his ankle.[7] The fact that A could have done otherwise than x does not entail that A could have done otherwise than try to x. Thus we can credit A with the range of abilities he eventually grows to have at the time he comes to have them without being forced to attribute that same range of abilities to him for all previous times. In short, our account does not render us incapable of explaining growth and acquisition of abilities.

The explanation of growth is of the utmost importance. For if it were impossible to explain the generation of wider-ranging abilities from narrower ones the hypothesis of beginninglessness would avail us little in our search for a solution of the freedom-determinism puzzle. It is this kind of change, the kind Bhattacharyya calls the "channelizing" of our normal activities, which leads on to increased skills, capacities, and abilities; in short which provides us with freedom. But if it is the case that in order that an agent have such increased abilities at any time he must always have had them, then surely growth is precluded. This is apparently a significant option in Indian philosophy. The Advaita Vedānta, for example, holds that the self has always been free, and in Buddhism Nāgārjuna seems to have held that there is no growth; in both systems enlightenment is sudden and does not involve progressive, causative growth toward freedom.

The formula covering the causes of actions which we might or might not perform is this: presence of opportunity for a suitably wide range of actions

[7] This rather unedifying example should not blind us to the fact that our ability to fail when we try to do something and instead do something else is a very important element in creativity. If we could never fail in what we try to do, we should not discover modes of action which prepare the way for new and broader abilities. A man who turns his ankle by mistake may learn to turn it on purpose if the occasion subsequently demands.

+ a suitably wide range of ability + trying. If each of these three causal factors is in turn caused, and each of the causal factors for those causes is caused, etc., then the resulting system is deterministic.[8] Indeed, such an account involves a harder determinism than is seemingly implied in Bhattacharyya's words quoted at the outset, for he apparently reserves certain events—those that proceed from detachment—as not fully caused.

But the determinism I have described does not preclude freedom, if by freedom we mean the ability to channel our actions in the directions we intend, coupled with the capacity for growth, for acquiring new abilities. For instance, take the supposed defect in any such deterministic account as ours, namely, that trying to x, itself an action, must have a cause and so we could not have helped trying to x. This objection falls to the ground provided we can explain how it is that at some point in the regress—x-ing, trying to x, trying to do that, etc.—there can be the production of a determinable ability from a determinate one, that is, of the ability-to-y-or-$\bar{y}$ from a mere ability to z. If we can explain this relation, then the fact that, say, we could not have done otherwise than try to x does not entail that we could not have done otherwise than x. It is the assumption of this entailment which is at the root of the particular puzzle about freedom and determinism now under consideration, the assumption, to repeat, that if A could not have done otherwise than try to x, then he could not have done otherwise than x.

The fault in the assumption in question might be labeled the Fallacy of Misplaced Determinateness. We are tempted to think that if A intends to do something, then there is some fully determinate event x he intends to bring about; thus either it is the case that he intends to bring about that event or it is not the case that he intends to bring it about, and that is the end of it. This is quite uncritical. When we try to do something, we are usually intending to bring about one or another of several events falling within a certain range. When we say "A tried to x and succeeded," we do not normally mean that A was fully cognizant in advance of the precise, determinate nature of the event which he in fact later produced, but rather that an event of its sort, perhaps not precisely that one but something enough like it to count, is the sort of event A wanted to produce, and he acted in such a way as happened to bring one such event about. It is normally not accurate to say that the agent intends to bring x about, if x is a fully determinant

[8] It is deterministic despite the fact that there may be occasions on which ability + opportunity + trying is not followed by success, since in talking of abilities we are inevitably involved in a statistical account. The statistical element in the analysis of "having an ability" proposed is, I think, an ineliminable aspect of any analysis of that phrase which will satisfy. This view is on the whole supported by Kaufman, *op. cit.*, and by Irving Thalberg, "Abilities and Ifs," *Analysis* 22 (1962), 121-126.

event. Rather we ought to say that the agent intends to bring about x or y or z, and performs action a by way of trying. In such an instance, then, A neither intends to bring about x nor does he not intend to bring it about.

When an agent tries to do something and succeeds, two events are fully determinate, namely, what he did when he tried and what he succeeds in doing. What is not determinate is what he tried to do, his intention. For example, A, trying to stamp his foot, flexes his leg muscles in a certain fashion and causes his foot to hit the ground at place p and time t with a determinate impact i. This determinate event is the action A succeeded in performing; the particular flexing of his leg muscles (at a certain time and place, etc.) is what he did in trying to stamp his foot; but what he tried to do was not to cause his foot to hit the ground at place p and time t with impact i, but to bring about either that determinate action or something enough like it to satisfy the description "stamp his foot" in "he is trying to stamp his foot." A given trying, then, is a determinate event considered under a determinable description. That such a determinate event is caused, that is, that the determinate flexing has a sufficient condition S, does not of itself guarantee that the determinate event constitutes a trying-to-stamp-his-foot action. Although given S the determinate flexing must follow, what need not follow is that that flexing be a case of A's trying to stamp his foot. By the same token, that A could not have done otherwise than try to stamp his foot does not of itself guarantee that the determinate flexing took place, and therefore since that determinate flexing is the cause of the determinate stamping at place p, time t, and with impact i, it might well have been that A, though he could not have done otherwise than try to stamp his foot, could have done otherwise than stamp it in that particular determinate way. And from that it follows too that he could have done otherwise than stamp it at all—as we have already suggested—even though he could not have helped trying to stamp it.

A trying, then, consists of a determinate event considered under a determinable description. This determinable description is produced in A by the nature of his abilities. Thus a trying is a result of conditions jointly sufficient to produce a determinate event—say, a determinate flexing—together with the conditions sufficient to produce in A the ability to intend a certain determinable range of events as outcomes of his actions.

If something is red, then it is some one determinate shade of red, say, crimson, and not scarlet or shocking pink or any other determinate shade of red. That is all right. But because A has an ability to x or $\bar{x}$, it does not follow that he has the ability to x, say, which precludes an ability to $\bar{x}$, even

though x and $\bar{x}$ are themselves mutually exclusive alternatives. A determinable ability, unlike a determinable shade of color, can occur without it being the case that only one of the mutually exclusive determinates falling under it occurs. (That is not to say that such an ability may occur without any determinate event occurring along with it: one may admit, without denying what is being maintained here, that at every moment in time there is at least one completely determinate event occurring.) Thus I may intend to x-or-$\bar{x}$ without it being either the case that I intend to x or that I intend to $\bar{x}$. Furthermore, I may try to develop the ability to x-or-$\bar{x}$ without either trying to develop the ability to x or trying to develop the ability to $\bar{x}$. We have seen how such determinable abilities enter into the causal account of actions. There is no reason to deny that one such ability may be caused to develop from others; we know that it happens. Abilities can be viewed as selective mechanisms, distinguishing a range of events according to a disjunction of relevant features. Intentional selectivity of this sort need involve no objectionable appeal to privileged access: abilities may not involve inner experiences at all in their identification, but be recognized and classified according to the characteristic behavioral symptoms displayed by those who have them. "Conceptual schemes" or "language-games" are larger-scale examples of abilities which function as selective mechanisms.

The point, however, is that the occurrences of such abilities are among the causal factors of human actions and can themselves be developed from more determinate, less wide-ranging abilities. It is this process to which I believe Bhattacharyya is referring when he speaks of the channelizing function of moral and spiritual activities. It is not that such selective sets or dispositions are uncaused, but rather that they are determinable without being determinate; thus their development can take place without having to be exercised in some determinate fashion upon their maturation. This corresponds to the experience of choosing, which many believe to be at the heart of the notion of freedom. Blind submission to emotions involves the production of relatively determinate actions by relatively determinate causes. Where trying comes into the picture, there enters an element of selection and choice. Freedom consists, according to the Indian view, in enriching the scope of one's abilities so that one is minimally dependent upon determinate passions produced by "external" events not of one's choosing. The detachment recommended in Indian texts is precisely this broadening of one's responsive awareness coupled with development of non-dependence on external passions.

If this is correct, we need not qualify the Indian commitment to deter-

minism to the extent that Bhattacharyya seems willing to. We may admit that our abilities are causal factors of our actions, and that every event—actions and possession of abilities included—has a sufficient condition. We have seen that it is possible, even under this form of determinism, to make sense of "could have done otherwise," of choice, growth, and acquisition of abilities, even of the characteristics of the Indian's complete freedom.

Despite all this, it is possible to raise doubts about freedom and determinism on grounds that are not touched by this account, along the following lines. Suppose that each *determinate* event which occurs in history, whether it is a human action, a trying, or however we choose to describe it, has as a sufficient condition other completely determinate events which in turn have their determinate antecedents, etc. If this is so, it is in principle possible for someone who knows all the laws of nature and who also knows the determinate state of affairs which pertains at a given point in time to predict all subsequent determinate events. Furthermore, this will be possible even if the universe is beginningless. Now since it is possible to predict each and every determinate event in this way, one might argue that determinism precludes freedom precisely because it allows for complete predictability.

This particular line of doubt is, to my knowledge, not a source of concern to Indian philosophers and I believe it is possible to understand why. Where the argument from predictability assumes that complete predictability undermines freedom, the Indian makes no such assumption. Since the freedom he dreams of is freedom from dependence on passions coupled with responsive awareness, he can claim with perfect justice that complete predictability, far from undermining freedom, provides conditions for it which, though perhaps not sufficient, are important and possibly even necessary. If part of one's aim in life is broadening one's responsive awareness, the more one knows of the determinate events which will occur, the more aware and the more responsive one can be. For example, if one possesses special powers of prediction and is at the same time responsive to the needs of others, one is in a position to assist them in improving their abilities. Recognition of this is characteristic of the Indian attitude toward the sage and the *guru*.

But surely if one could predict with certainty each and every event one would not try to do anything! Though this sounds like a terrible thing to us, the Indians welcome it and endorse it as precisely what is to be worked for. After all, striving is the condition of bondage and results from dependence on our attachment to things, and on passions whose source is external to us. If we were indeed fully omniscient we would, as has just been said, not try to do anything, that is, to seek anything. If this raises

any problems at all for Indians, it raises them for a view such as Mahāyāna Buddhism's doctrine of a *bodhisattva* who is supposed to be omniscient and still capable of working to help others achieve freedom. One may well wonder whether these attributes are consistent, and the issue is to some re-flected also in the Hindu debate over *jīvanmukti* and *videhamukti,* whether one can be free and still living.

A full treatment of this last point lies beyond the scope of this paper. We may ask, finally, whether the possibility of attaining complete predictability militates against anything that was offered concerning the nature of the relative freedom one gets through broadening one's abilities. I think the answer is no. Who does the attaining? If I attain this kind of omniscience, I shall have a condition of complete freedom in my grasp, but this need not affect you. If I offer you my knowledge of what will happen, as a way of assisting you to your freedom, you may not believe me, you may not under-stand me, you may not know how to use the information I give you. Unless you too are free you will still have to grow in your abilities, responsiveness, and control. Indeed, so may I, even in my omniscience, though it seems dif-ficult to deny that ability to foresee the future must give a man considerable advantage in developing his attitudes commensurately with his intellectual prowess, provided he has the wit to use it. But one man's foreknowledge is not necessarily pertinent to another's problems, and for our problems in this world, lacking foreknowledge, the utility of the categories of abilities, oppor-tunity, and trying is unaffected by the possibility that we may eventually gain omniscience. Man may collectively attain foreknowledge through sci-ence; there is no guarantee that this will provide him, individually, freedom of the sort the Indian works for, though under certain conditions it might go a long way toward that goal.

What is important is just to see that the possibility of complete predicta-bility solves no actual problems which face man in the conduct of his affairs, except insofar as it stands as an ideal for rational inquiry to aspire to. The problem of gaining freedom is one we cannot merely put aside until science is able to predict all, and that is why we must adopt categories for the analysis of human causality and freedom other than those limited to rela-tions among fully determinate events. Freedom as a practical goal does in-deed require indeterminacy, but it does not require indeterminism, and it cer-tainly does not require abandonment of the scientific effort to relate events in their ultimate determinateness. It does, however, require us to discuss the causes of our actions in terms which we can understand and use now, and given our current powers, these terms must involve indeterminacy. Perhaps

we shall never have complete knowledge of the laws governing events in all their determinacy; perhaps we can never hope to know the completely determinate state of the universe and so will not in fact ever be able fully to predict the future. Perhaps, too, we shall never become completely free, i.e., responsively aware and yet unattached to our passions. That is no reason not to work for both goals, and there is no incompatibility between them.

JAMES DUERLINGER

REDUCTIONIST AND NONREDUCTIONIST THEORIES OF PERSONS IN INDIAN BUDDHIST PHILOSOPHY

INTRODUCTION

In *Reasons and Persons*, Derek Parfit quoted passages from Th. Stcherbatsky's English translation of the ninth chapter of Vasubandhu's *Treasury of Knowledge* (*Abhidharmakośa*), a 5th century Buddhist text, in support of his claim that the adoption of a reductionist theory of persons is not a peculiarly Western phenomenon.[1] This chapter has been recently retranslated into English, under the title, 'Refutation of the Theory of Selfhood: a Resolution of the Questions About Persons' specifically for the purpose of facilitating its careful study by Western philosophers.[2] This new translation, which I shall call the 'Refutation,' does indeed help us to see that the chapter contains what may be regarded as the classic statement of the Buddhist version of the reductionist theory of persons of the sort held by David Hume and Parfit himself.[3] In what follows I shall show that Vasubandhu's statement is in fact a statement of the reductionist theory and explain in some detail the form this theory takes. Moreover, since the scripture-based Sautrāntika theory set out in Vasubandhu's statement is virtually the same as the Vaibhāṣika theory (as well as the Theravāda theory), and is one with the theories held by the Sautrāntika logicians, Yogacārins, and Svātantrika-Mādhyamikas in its reductionist character, I shall use it here to illustrate the reductionist character of the theory of persons attributed to the Buddha within most of the Indian Buddhist schools of philosophy.[4]

But reductionism is not, as Parfit and many others have assumed, the only credible and widely-accepted interpretation of the Buddha's theory presented by the Indian Buddhists themselves. In the 'Refutation' itself, of course, Vasubandhu presents and rejects the Vātsīputrīyas' nonreductionist account of the Buddha's theory. But their account, which became a favorite target of the Indian Buddhist

philosophers,[5] is not the form the nonreductionist theory ultimately takes within Indian Buddhism. Here I shall only comment on the general form of their interpretation. My chief concern will be a much more successful nonreductionist challenge to the reductionist interpretation of the Buddha's theory. This is the challenge made in the 8th century A.D. by Candrakīrti (hereafter, Candra), the founder of the Prāsaṅgika interpretation of Nāgārjuna's Mādhyamika school of Buddhist thought, in his *Introduction to the Middle Way: Verses with Commentary*.[6] On the basis of both independent argument and scriptural authority Candra presented a refutation of the reductionist interpretation and carefully formulated a Buddhist version of the nonreductionist theory of persons which accorded with the principles of the Mādhyamika philosophy set out by Nāgārjuna. That we may appreciate how, within the Indian Buddhist tradition, both reductionist and nonreductionist theories of persons were accommodated, I shall also explain Candra's nonreductionist theory and his rejection of Buddhist versions of the reductionist theory.[7]

To provide a classification of theories of persons adequate to the completion of my task I shall begin with the assumption that Indian theories of persons, unlike Western theories, are predicated on the view that persons, as actually conceived in our everyday life from the first person singular perspective, do not really exist in the sense that they do not, so conceived, exist independently with natures of their own on the basis of which they are named and conceptualized.[8] Consequently, Indian theories of persons are basically attempts to explain what our true nature is and in what way we do in fact exist. The basic problem of self, in this context, is that our actual first person singular concept of ourselves, when subjected to careful analysis, can be shown to be without a real object, since it contains metaphysically irreconcilable components. For our purposes, we may say that the problem is generated by the fact that this concept contains both reductionist and nonreductionist components as they are defined by Parfit (and explained below). For instance, if we analyze how we actually talk and think about ourselves in everyday life from the first person singular perspective, we shall discover that we (i) appear to be distinct from our bodies and minds as unitary possessors of them, yet also to be the same as our bodies and minds by virtue of possessing

both physical and mental properties, and (ii) appear both to be and not to be the same over time in dependence upon appearing to be both distinct from, and the same as, our bodies and minds. That nothing could answer to such a self-contradictory description is clear enough. Theories of persons, insofar as they are formulated in response to this problem, are attempts to deal with the fact that our actual concept of ourselves would seem to be without a real object.

Since we cannot very well simply abandon a first person singular concept of ourselves, the Indian philosophers respond to this problem in one of two ways. The most common response to the problem is to attempt to revise our actual concept of ourselves in an effort to provide it with a real object of some sort. So it is argued that this or that component of the concept is to be excised, and that persons, correctly understood, really exist in some way. All such theories of persons may be called revisionist.[9] All reductionist theories of persons are revisionist in character because they excise the nonreductionist components of our actual concept of ourselves. Nonreductionist theories of persons will be revisionist in character when they excise the reductionist component of the concept.

The less common response to the basic problem of self is to retain our actual concept of ourselves and attempt to explain in what way we exist as we actually conceive ourselves. Some nonrevisionists argue that persons really exist, but in an inexplicable or indefinable form, as the Vātsīputrīyas do, and others outright reject the real existence of persons, as Candra does, and claim that persons exist in name or concept only, or merely by convention, in the sense that, although they do not by themselves possess a nature by virtue of which they can be named and conceptualized, they are assumed, by convention, to possess such a nature. The former view may be called the realist, the latter, the conventionalist, version of the nonrevisionist theory.[10] All nonrevisionist theories are nonreductionist in character.

The reductionist theory, as Parfit presents it, includes theses about both our existence and our identity over time. The theses are (i) that our existence is reducible to the existence of a brain or body and the occurrence of a series of interrelated physical and mental events, and (ii) that our identity over time is reducible to a set of impersonal facts about a brain or body and the occurrence of a series of interrelated

physical and mental events. Impersonal facts, he explains, are those which can be described without presupposing our existence or identity and without explicitly claiming that we have experiences. Parfit says that reductionists claim that we exist, but need not explicitly claim that we exist, since in their view a complete description of reality can be given without claiming that we exist. What is common to every nonreductionist theory, as characterized by Parfit, is simply the denial of the above two theses.[11]

Parfit's characterization of the reductionist view of our existence implies an implicit reductionist thesis concerning our unity. For if our existence is reducible to that of our bodies or brains, which are made up of parts, and a series of physical and mental events, more than one of which can occur at any given moment in the series, our unity at any given moment must also be reducible to a relation between our bodily parts and the collection of our physical and mental states at that moment. In what follows I shall take it for granted that the reductionist theory of our existence has this further reductionist thesis as an implication and that the nonreductionist denies it. Moreover, since the impersonal facts to which Parfit would reduce our identity at different times would seem to be facts about a continuum of physical and mental states, activities, or processes bound together by relationships which preserve the character of the phenomena within it on the basis of which our actual concept of ourselves is formed, in what follows I shall more simply speak of the reductionist and nonreductionist theories as the views that our existence and identity at different times are and are not, respectively, reducible to those of our bodies and mental states in a continuum which preserves the character they possess in reliance upon which our actual concept of ourselves is formed.

The classical form of the nonreductionist theory is the separate self theory, according to which we possess an existence and identity over time separate from those of our bodies and mental states in a continuum. Vasubandhu explicitly challenges versions of the separate self view as presented by the Sāṃkhyas and Vaiśeṣikas, whom in the 'Refutation' he calls the Tīrthikas. The Vātsīputrīyas propound a minimalist version of the nonreductionist theory which Vasubandhu in the 'Refutation' subjects to an extensive analysis. This is the view that

our existence and identity over time are neither reducible to, nor separate from, those of our bodies and mental states in a continuum. The Vātsīputrīyas claim that we really exist, in spite of the inexplicability of our existence. Candra rejects the Vātsīputrīyas' claim on the ground that the inexplicability of our existence is the very reason our real existence is to be denied.[12] Vasubandhu and Candra agree that our actual concept of ourselves appears to have, but does not have, a real object, while the Vātsīputrīyas claim that its object is real, but falsely appears to exist separately from our bodies and mental states. Candra does not, as Vasubandhu does, attempt to revise the concept for the sake of providing it with a real object, nor does he, as the Vātsīputrīyas do, provide it with a real object whose existence is inexplicable.

Since I have included Vasubandhu among those who hold a reductionist theory of persons, even though he does not explicitly say that our existence and identity over time are reducible to those possessed by our bodies and mental states in a continuum, I must explain why I have classified his theory in this way. First of all, we should notice that Parfit's characterizations of reductionism and nonreductionism do not commit their adherents to any particular metaphysical analysis of our bodies and mental states. Parfit himself states, in fact, that physicalist, mentalist, and dualist versions of the reductionist theory are possible.[13] For this reason these versions may be classified as different forms of the reductionist theory. Secondly, Parfit's characterizations do not commit reductionists or nonreductionists to any particular theory either about the ontological status of the body and mental states or about the nature of the continuum of the body and mental states. But such differences can be used to distinguish different versions of the physicalist, mentalist, and dualist forms of the reductionist theory of persons. The more important of the Indian philosophers agree that the continuum is causal in nature, but not all of these agree about the nature of the causal continuum. Among the Buddhists, the Vaibhāṣikas, Sautrāntikas, and Yogacārins assert that the causal continuum is made up of phenomena which exist independently with natures of their own. The Svātantrika-Mādhyamikans deny that its constituents exist independently, but assert that they possess natures of their own by reason of appearing to the conscious-

nesses in dependence upon which they exist. Let us call these, respectively, the realist and modified realist versions of reductionism.[14] The distinction between realist and modified realist versions of reductionism, however, is rejected by Candra. He claims that the idea of a phenomenon which possesses a nature of its own by reason of appearing to the consciousness in dependence upon which it exists is incoherent, since if a phenomenon possesses a nature of its own, its nature cannot be possessed by reason of appearing to a consciousness. Consequently, Candra intends his criticisms of realist versions of reductionist to apply also to modified realist versions.[15]

It follows that a theory about our **existence** can take a variety of forms. In particular, it can take the form of Vasubandhu's theory, that we are collections of the five aggregates (*skandha*-s) into which he believes our bodies and mental states are analyzable. Because the first of the five aggregates, according to Vasubandhu, is physical in nature, while the remainder are mental, he holds a dualist form of the reductionist theory. Moreover, since he believes that the aggregates are substantially real (*dravyasat*), his theory is realist in character. Vasubandhu's version of the reductionist theory about the unity of persons is that we are one not because we possess a single nature on the basis of which we are named and conceptualized, but because our unity consists in the fact that our aggregates co-dependently arise at the same time. We are one, he believes, in name and concept only.[16]

Parfit's characterization of reductionist and nonreductionist views of **personal identity** can also be applied to views held by Vasubandhu and his nonreductionist opponents. Vasubandhu does not specifically speak of personal identity in his verbal confrontations with the Vātsīputrīyas and the Tīrthikas and his views will not be familiar to most Western philosophers, but Parfit's characterizations do apply to views held by them. For Vasubandhu claims that we are not, as the Tīrthikas' believe, permanent (*nitya*) in the sense that we exist over time without change, or impermanent (*anitya*) in the strict sense that we exist only for a moment, but rather exist over time as a causal continuum of impermanent phenomena which retain the characteristics by virtue of which we are named and conceptualized.[11] The notions of a permanent phenomenon and of a causal continuum of impermanent phenomena surely are, at least in part, competing metaphysical explanations of our ascriptions of identity over time to ourselves. In

fact, the claims of Vasubandhu and the Tīrthikas may be construed, respectively, as metaphysical renditions of what many Western philosophers have called loose or imperfect identity over time and strict or perfect identity over time, and of what Parfit takes, in relation to persons, to be identity over time as a fact reducible to a set of impersonal facts and identity over time as an irreducible fact about a separately existing entity.

The minimalist nonreductionists hold the view that our identity over time cannot be explicated in either of these two ways. But to the extent that they deny that our identity over time can be reduced to that of a continuum of impermanent phenomena, their theory may be counted as nonreductionist in character. Thus, since Vasubandhu claims that we are the five aggregates as a collection *in a causal continuum*, his view may be counted as a version of the reductionist view of personal identity. Both the separate self view, that we are permanent phenomena, as well as the minimalist view, that our identity over time cannot be explicated in either of these ways, may also be counted as nonreductionist views.

VASUBANDHU'S STATEMENT OF HIS THEORY OF PERSONS

Seeing Oneself as a Self and the Cause of Suffering

Vasubandhu begins the 'Refutation' by claiming that the practice of the theory of liberation he has just finished systematically presenting in verse and refining in his commentary is the only means by which we can gain freedom from suffering, since the practice of the rival theory of the Tīrthikas commits their practitioners to seeing themselves as selves (*ātmadṛṣṭi*), which is the root cause of the mental afflictions, which by contaminating actions produce all suffering.

There is no (way to achieve) liberation (from suffering) other than (through the practice of) this (teaching of the Buddha), since (in the teachings of the Tīrthikas the error of) seeing (oneself as) a self is to be found.[18] (Those who practice their teachings will not be liberated from suffering,) for they do not understand that the name and concept of a self[19] is applicable only to (a collection of) aggregates in a (causal) continuum. They believe, (instead,) that a self is a substance that is separate (in existence from any and all of the aggregates). (But) the mental afflictions, (which cause all suffering,) arise from clinging to (oneself as) a self.

Vasubandhu does not think that we see ourselves as selves because we adopt the Tīrthikas' separate self theory of persons. In this opening passage of the 'Refutation' he is claiming that the practice of the Tīrthika theories of liberation will *reinforce* our inborn disposition to see ourselves from a nonreductionist perspective. From his point of view, their theories of persons are attempts to provide a metaphysical foundation for the nonreductionist components of our actual concept of ourselves, while excising its reductionist components. The principal Tīrthika theories of liberation to which Vasubandhu objects in the 'Refutation' are those of the Sāṃkhyas and the Vaiśeṣikas, who claim (i) that we are *substances* in the sense that we exist independently with single natures of our own, and so exist independently of the ever-changing multiform bodies and mental states we are said to possess, and (ii) that we are, by virtue of the natures we possess, *partless* in the sense that we are indivisibly one at any given moment and *permanent* in the sense that we exist over time without change. The Tīrthikas claim that we bring about our suffering by falsely identifying or associating ourselves with our bodies and/or mental states, but can free ourselves from suffering by learning to identify ourselves with the substances we really are. A person of this sort they call a self (*ātman*) and argue that it can be known to exist by sound inference. Their theories of persons as selves differ in other respects, in the Vaiśeṣika theory, for instance, agency being assigned to a self, and in the Sāṃkhya theory, agency being denied. But the theories of both, like that of Vasubandhu, are revisionist to the extent that they are meant to correct our actual concept of ourselves, which is believed to incorporate errors concerning our true nature the acceptance of which causes suffering.

Although the other Indian Buddhist philosophers about whom we have information claim that we suffer because we see ourselves as selves, not all agree about what it is to see ourselves as selves, and consequently, about what the ultimate cause of suffering is. The standard Buddhist characterization of seeing ourselves as selves, that we cling to the existence of an 'I' (*aham*) and a 'mine' (*mama, ātmīya*), makes it clear that the error is that we are attached to the existence of that about which we are talking and thinking when we talk or think about ourselves from the first person singular perspec-

tive. Vasubandhu's own view is that our attachment is to an irreducible existence which we appear to have when we make reference to ourselves from the first person singular perspective. In dependence upon the appearance of ourselves as possessed of an existence not reducible to the existence of our aggregates, Vasubandhu believes, we also appear to ourselves both to be irreducibly one at any given moment and to be irreducibly the same over time. This further appearance encompasses the remainder of the nonreductionist component of our actual concept of ourselves as persons. However, this appearance is not the one whose acceptance Vasubandhu calls seeing oneself as a self and its acceptance is not the root cause of suffering. But since this second appearance is a consequence of the first, Vasubandhu often in the 'Refutation' attempts, within the context of replying to nonreductionist objections to his theory, to explain it away within his own reductionist framework.

Vasubandhu's theory of why we suffer is criticized by Candra, who claims (i) that we do not really exist, since our existence cannot be separate from, or reducible to, that of the phenomena in reliance upon which we are named and conceptualized, and (ii) that we suffer because we see ourselves as selves in the sense that we see ourselves as phenomena that really exist.[20] He denies that we suffer because we cling to an existence for ourselves which is not reducible to the existence of our aggregates as a collection, since in his view our existence is not in fact so reducible. He argues that because we exist *in dependence upon* the existence of our aggregates, our existence cannot be reduced to that of our aggregates.[21] Vasubandhu and the other Buddhists who embrace a realist or modified realist form of reductionism fail to explain why we suffer, Candra concludes, because, even though they claim that we are real in name and concept only, they commit themselves to our real existence when they reduce our existence to that of the phenomena on the basis of which we are named and conceptualized and claim that these phenomena really exist.[22] Candra maintains that the adoption and meditative practice of Buddhist versions of the reductionist theory of persons cannot be the means by which we eliminate our suffering, since what it is about our actual concept of ourselves attachment to which causes us to suffer is not properly identified.[23]

Vasubandhu's Basic Argument For This Theory

Vasubandhu's claim that those who would achieve liberation from suffering should accept his theory is immediately followed by another claim, which if translated literally, may be rendered:

> The expression, 'self' (*ātman*), is known to be applicable to the aggregates in a continuum and to nothing else because there is no direct perception or sound inference.

Although he says in this statement that the expression, "self," is known to be applicable to the aggregates in a continuum and to nothing else, Vasubandhu normally denies that a self exists. The context of the passage, moreover, suggests that Vasubandhu uses "self" in the statement only because he wishes to say that the Tīrthikas' own word for a person, "self," if used as a name for a person that really exists, must be the name of the aggregates as a continuum. Accordingly, I shall hereafter follow Vasubandhu's more customary practice of using the term, "self," to express a nonreductionist conception of persons and restate his own reductionist thesis about persons by saying that persons are selfless. The first part of Vasubandhu's statement, therefore, may be recast as the statement that the term, "person," is known to be applicable to the aggregates and not to a self.

Since the continuum of the aggregates about which Vasubandhu speaks he believes to be a *causal* continuum, and strictly speaking, he means to be speaking here of the aggregates *as a collection*, the first part of his statement quoted above may be expressed as the view that the term, "person," is known to be applicable to a collection of aggregates in a causal continuum and not to a self. Finally, this part of Vasubandhu's statement may also be freed of its exclusive linguistic trappings if we make use of the locution, "name and concept" (*prajñapti*),[24] which Vasubandhu usually employs in statements of his theory, in place of the locution, "the term, 'person.'" In this way we can more readily see that included within the scope of his statement is our use of the first person singular personal pronoun to refer to ourselves. He is saying, then, that the name and concept of a person are known to be applicable to the aggregates as a collection in a causal continuum and not to a self.

The reason they are applicable to the aggregates as a collection in a causal continuum, according to Vasubandhu, is that the aggregates as a collection in a causal continuum are the phenomena in reliance upon which a person is named and conceptualized. Consequently, he is assuming that the name and concept of a person must be applicable to the phenomena in reliance upon which the person is named and conceptualized. In other words, he assumes that our existence is reducible to that of the phenomena in reliance upon which we are named and conceptualized.

The Vaibhāṣikas, Sautrāntikas, Yogacāra logicians, and Svātantrika-Mādhyamikas agree that one or more of the five aggregates as a collection in a causal continuum is or are the phenomena in reliance upon which we are named and conceptualized.[25] Among these, the Vaibhāṣikas and Vasubandhu's Sautrāntika followers agree that the phenomena on the basis of which we are named and conceptualized are all five aggregates; the others claim that the bases are causally connected subtle consciousnesses which fall within the aggregate of consciousness. The scripture-based Yogacārins claim that they are consciousnesses which fall outside that aggregate and are collectively called the foundation consciousness (*ālayavijñāna*). All but the Prāsaṅgika-Mādhyamikas claim that we must exist among the phenomena in reliance upon which we are named and conceptualized or we do not exist at all, not even in name and concept only. Candra, however, denies that we are in fact named and conceptualized, since analysis shows that we do not exist independently with natures of our own on the basis of which we can be named and conceptualized.

We must now deal with the question of how exactly we are to understand what Vasubandhu means by claiming that the absence of direct perception and sound inference establish that the name and concept of a person are applicable to a collection of aggregates in a causal continuum and not to a self. A first step in answering this question is to notice that in claiming that direct perception and sound inference *cannot* be used to show that the name and concept of a person are applicable to a self, Vasubandhu surely means to be implying that direct perception and sound inference *can* be used to prove that they are applicable to the aggregates in a causal continuum. To explain how in these ways Vasubandhu thinks that the presence

and the absence of these sources of knowledge are relevant to the determination of that to which the name and concept of a person are applicable, we need to make two assumptions: first of all, that there are phenomena in reliance upon which persons are named and conceptualized, and secondly, that if persons really exist, they exist among these phenomena. The roles played by the presence and absence of direct perception and/or sound inference will then be to establish, respectively, that these phenomena are a collection of aggregates in a causal continuum, and that there is no self among them.

Consequently, Vasubandhu's statement involves the use of a two-part argument, one directed against nonreductionist theories of persons, and the other advanced in support of his own reductionist theory. The combined argument to which Vasubandhu alludes I shall call *Vasubandhu's basic argument.* Of the two arguments which comprise the basic argument, the first is to the effect that the name and concept of a person are known not to be applicable to a self because a self is not known by direct perception or sound inference to exist among the phenomena on the basis of which a person is named and conceptualized. I shall call this his *selfless persons argument* (hereafter, SPA). The purpose of the argument is to bring us to the realization that, even though we appear to be selves, we are not in fact selves, since a self is not known to exist among the phenomena in reliance upon which we refer to ourselves from the first person singular perspective. The second of the two arguments is that the name and concept of a person are known to be applicable to a collection of aggregates in a causal continuum because the phenomena on the basis of which the person is named and conceptualized are known by direct perception or sound inference to be a collection of aggregates as a causal continuum. I shall call this his *continuum persons argument* (hereafter, CPA). The purpose of this argument is to enable us to see that, since it is known that there is a continuum of aggregates in reliance upon which we refer to ourselves from the first person singular perspective, there really is something to which we refer when we refer to ourselves, even if it is not a self.

(i) *The Selfless Persons Argument*

The SPA is clear enough from our familiarity with Hume's famous argument in the *Treatise of Human Nature* against the existence of an irreducible self.[26] In saying that the idea of a self cannot be traced to a simple impression, Hume was in effect saying that he could find no directly perceived phenomenon on the basis of which he names and conceptualizes himself as a self. Hume, like Vasubandhu, also thought that the absence of a sound inference for the existence of a self plays a part in his denial of its existence. For he argued that there is no proof that a self is needed to explain the occurrence of the perceptions on the basis of which persons are named and conceptualized.[27]

In the SPA it is assumed that what is not known to exist by means of direct perception or sound inference is known not to exist. But, strictly speaking, the fact that a self is not directly perceived and that there is no sound inference to its existence warrants only the conclusion that a self is not known to exist, not that it is known not to exist. In the 'Refutation' Vasubandhu immediately follows his basic argument for his theory with an attempt to defend this assumption. In support of the claim that what is not known to exist is known not to exist, he argues that the twelve bases of perception (*āyatana-s*), which comprise all real phenomena, are known to exist by direct perception or sound inference. However, in making the assumption that the twelve bases of perception are an exhaustive classification of all real phenomena, he neglects to tell us how this is known to be true. His view, it seems, is that it is known to be true because the Buddha, an omniscient person, has said that the twelve bases of perception are an exhaustive classification of all truly existent phenomena. Hence, the problem of justifying the assumption that whatever is not known to exist is known not to exist is simply shifted to the equally difficult problem of justifying the claim that the Buddha is an omniscient person. In the absence of a demonstration that what is not known to exist is known not to exist, it would seem that we need not accept the SPA as it stands.[28]

All of the Indian Buddhist philosophers, except perhaps for Candra and his Prāsaṅgika followers, use the SPA to prove that there is no self. The proof is used to begin the process of destroying our seeing of

ourselves as selves by transforming, through the practice of concentration, our inferential realization that we are selfless into a direct yogic perception of our selflessness. It is used by them, however, only to the extent that "self" is defined in accord with their own theories of persons. Vasubandhu and most others use "self" in the SPA to refer to the nonreductionist component of our actual concept of ourselves. But the Vātsīputrīyas use it to refer to a separately existing self. Candra, to my knowledge, does not use the argument. His own basic argument, as I said above, is that we are selfless in the sense that we do not really exist, since our existence is neither separate from, nor reducible to, that of the phenomena in dependence upon which we are named and conceptualized.

(ii) *The Continuum Persons Argument*

The key to understanding the CPA is to see in some detail exactly how Vasubandhu uses inference to reduce our existence to the existence of our aggregates. In the 'Refutation' an inference of an analogous sort occurs[29] where he argues that the existence of milk is reducible to the existence of the collection of all of its elements. The inference, once fully reconstructed, is that since it is true (i) that milk really exists, (ii) that by direct perception we know that the constituents of milk really exist, (iii) that milk cannot exist apart from its constituents, (iv) that milk cannot exist as some of its constituents to the exclusion of the others, and (v) that milk cannot exist as all of its constituents individually considered, it follows (vi) that milk must exist as all of its constituents as a collection.

This inference can be extended to accommodate the view that milk really exists as all of its constituents collectively in a causal continuum. For the efficacy of the convention that milk is the same over time can only be explained, he believes, by the supposition that its constituents as a collection enter into a causal continuum of the sort which enables the convention to be useful for the practical purposes of life. If its constituents as a collection did not enter into such a continuum, it would not even be the same over time in name and concept only.

For like reasons, Vasubandhu also believes that a person is known to be all of its constituents collectively in a causal continuum. We need not fully reconstruct this part of the argument in order to see

that Vasubandhu attempts to provide the convention that we are the same over time with a metaphysical foundation. But he does not, as Hume does, also attempt a detailed account of how it comes about that we adopt this convention. He simply assumes that our appearance to ourselves of being irreducibly the same over time is an immediate consequence of our more fundamental appearance to ourselves of being irreducibly existent and one.

It would seem that among the Indian Buddhists only Vasubandhu and his followers accept the CPA as it stands. The Vātsīputrīyas would explicitly deny that if we really exist, we must exist among the phenomena on the basis of which we are named and conceptualized. Although Candra would not deny this, he denies that we really exist.

AVOIDANCE OF THE TWO EXTREMES AND THE MODE OF EXISTENCE OF PERSONS

Together the SPA and CPA arguments are used by Vasubandhu in order to safeguard what the Indian Buddhist philosophers call the middle way between the extremes of eternalism and nihilism. Extreme views concerning our existence are distinguished from extreme views concerning our identity over time. According to Vasubandhu, eternalism with respect to our existence is the view that we are substantially real and nihilism is the view that we are not real at all. The Tīrthikas clearly fall to the eternalist extreme concerning our existence and the Vātsīputrīyas fall to it, Vasubandhu believes, in a subtler form, since they deny that we are real in name and concept only. In the 'Refutation' he also says, in a single sentence, that the Mādhyamikas fall to the nihilist extreme because they deny the real existence of all phenomena.[30] With respect to personal identity, he believes, eternalism is the view that we are permanent in the sense of existing over time without change, and nihilism is the view that we are impermanent in the sense of existing only for a moment. The Tīrthikas clearly fall to the eternalist extreme concerning personal identity, and the Vātsīputrīyas, he thinks, are committed to it by their denial that we are the same over time in name and concept only. No one, apparently, holds a nihilist view of personal identity of the sort Vasubandhu rejects.

Vasubandhu tries to avoid *eternalist* views of our existence and identity over time by the use of the SPA. We cannot be substantially real because we are not selves; and since we are not substantially real, and only if we were substantially real could we be permanent, we are not permanent either. He tries to avoid *nihilist* views of both our existence and identity over time by the use of the CPA. Although we are not substantially real, we do possess a nominal and conceptual reality in dependence upon our existence being reducible to the existence of the collection of aggregates; and although we are not the same over time in the way a permanent substance is, we are not impermanent either, since we possess a nominal and conceptual identity over time in dependence upon the causal connectedness and continuity of the collection of our aggregates.

It was to avoid the extremes of eternalism and nihilism that the Buddha himself distinguished modes of existence a phenomenon known to exist may possess and classified all things known to exist by reference to whether they possessed one or the other of them. A phenomenon classified from the point of view of its mode of existence he called a reality (*satya*) and said that it is real (*sat*), rather than not existing at all or being unreal (*asat*). Accordingly, when the Buddha introduced a distinction between conventional deceptive realities and realities which he called ultimate,[31] he was distinguishing two modes of existence possessed by phenomena known to exist. Although the Buddha distinguished these kinds of realities he did not unambiguously explain what they are and how they are related to one another. Consequently, one of the central concerns of Indian Buddhist philosophers is to provide an unambiguous explanation. Their disagreements about this constitute one of the central themes of the history of Indian Buddhist philosophy.

In the *Treasury of Knowledge* Vasubandhu presents, and in his commentary on it apparently accepts, the Vaibhāṣika accounts of these two realities. The accounts are operational definitions in which we are given the means by which to determine whether a phenomenon is a deceptive conventional reality or an ultimate reality. A deceptive conventional reality is defined as a phenomenon whose nature can no longer be discriminated when it has been taken apart physically or mentally analyzed into constituent parts of different sorts. An ultimate

reality is defined as a phenomenon whose nature can still be discriminated when it has been taken apart physically or mentally analyzed into constituent parts of different sorts. Consequently, a deceptive conventional reality is deceptive insofar as it appears to have a nature it does not really possess, but instead is really a mere collection of the different sorts of constituents into which it can be analyzed, and it is conventional insofar as it is by convention named and conceptualized on the basis of the presence of its constituents. An ultimate reality is a phenomenon which does really possess the nature it appears to possess and is named and conceptualized on the basis of its possession of that nature by itself. These two realities are to be contrasted to unreal phenomena, which appear to possess natures of their own, but, since they cannot be found to exist among the phenomena on the basis of which they are named and conceptualized, do not exist at all.

In accord with these definitions Vasubandhu can say that a person is a deceptive conventional reality, that the most elemental phenomena of which the person is composed are ultimate realities, and that a self is not a reality at all. When persons, as they appear to a consciousness, are subjected to Vasubandhu's test of reality-kind, they are found to be deceptive conventional realities because their natures can no longer be discriminated when they are analyzed into the phenomena on the basis of which they are named and conceptualized: none of the phenomena into which they can be analyzed is found to possess the nature of a person. Therefore, though persons exist as phenomena which appear to possess single natures by virtue of which they can be named and conceptualized, they do not possess such natures, and are in fact real in name or concept only. Because we assume, in accord with the false appearance of ourselves, that we are single-natured phenomena whose existence is not reducible to that of our aggregates as a collection, we act on this assumption and thereby cause ourselves to suffer. Since the persons who falsely appear to be single natured and who make this assumption exist as deceptive conventional realities, the nihilist extreme is avoided.

Candra believes, however, that if persons, as deceptive conventional realities, are in fact the aggregates as a collection in a causal continuum, what we need to say about ourselves for the practical purposes of life we must also be able to say about the aggregates as a

collection in a causal continuum, since in this account every reference to ourselves must be a reference to them. However, we cannot say about the aggregates as a collection in a causal continuum what we say about ourselves, since they lack the unity we must possess if we are to possess bodies and mental states and to be subjects of different mental states at the same time, and they lack the identity over time we must possess if we are to be agents of actions which receive the results of those actions in accord with the law of actions and their results.[32] Candra rejects the attempt to ground our ascriptions of identity over time to ourselves in an account of a causal relation between the really existent aggregates in reliance upon which we are named and conceptualized. He argues, following Nāgārjuna,[33] that a causal relation cannot obtain between phenomena which really exist.[34] In any case, he adds, a collection of independently existent aggregates, since without real unity, cannot be an agent of action or subject of experiences, and hence, if our existence is reducible to that of a collection of aggregates, we must reject the conventions that we are agents of actions and subjects of experience.[35] In other words, the reductionist theory preserves our real existence at the expense of calling into question our most essential conventional beliefs about what we are.

Candra places great weight upon the preservation of conventional beliefs about what we are because the Buddha said that he does not dispute the truth of the conventions of the world.[36] In particular, Candra believes, the Buddha never intended that we *revise* our actual concept of ourselves, but that we *retain* it so that we can use it in an effort to realize our mode of the existence. The concept must also be retained by Buddhists, he believes, because it is needed in order to formulate the problem of suffering in terms of our acceptance of the false appearance of ourselves as really existent. The more general point, of course, is that it must be retained because, in spite of the fact that it cannot be used to refer to anything, all of its components, including its nonreductionist components, are needed if it is to be a useful tool for the conduct of human affairs. By rejecting our actual concept of ourselves, Candra also believes, Buddhist proponents of realist and modified realist forms of reductionism fall prey to the nihilist view that we do not exist at all as we actually conceive ourselves. And by revising the concept, he believes, they fall prey to

the eternalist view that we really exist, since even though they deny that we really exist as what we actually conceive ourselves to be, they assert that there is in fact something real to which we refer when we attempt to name and conceptualize ourselves.

What we need to realize about ourselves in order to become free from suffering, according to Candra, is our ultimate reality, which is our emptiness (*śūnyatā*) or absence of real existence. We are deceptive conventional realities in the sense that when we name and conceptualize ourselves, we falsely appear, to our ordinary consciousness, to exist independently with natures of our own by virtue of which we are named and conceptualized, when in fact we exist in dependence upon the functionally necessary convention of being named and conceptualized in reliance upon the presence of the aggregates as a collection in a causal continuum.

The downfall of all revisionist theories of persons, according to Candra, is that the elimination of either the reductionist or the nonreductionist components of our actual concept of ourselves has absurd consequences by reason of undermining our conventional idea of what we are. The separate self version of the nonreductionist theory, as Buddhists of all stripes insist, undermines the convention because it makes impossible, between the self and the causal continuum of one's body and mental states, a causal link on the basis of which we can be said to suffer and to be capable of freeing ourselves from suffering. The well-known Western analogue to this criticism is the objection to Descartes' separate self theory that it cannot explain how the mind can interact with the body. And the reductionist theory, as its nonreductionist opponents are fond of pointing out, undermines our conventional idea of what we are because it makes impossible, between ourselves at different times of our existence, a relation on the basis of which it becomes rational for us to seek the elimination of our suffering in the future, or, between different experiences we may have at one time, a relation which explains why they are all experiences belonging to one subject. In the West, Locke, Butler, Reid, Kant, and more recently, Chisholm, all argue this point in different ways and contexts.

It is in fact not just from an Indian philosophical perspective that the reductionist theory and the separate self form of the nonreduc-

tionist theory have been seen to be problematic because of their revisionist character. Part of the genius of Wittgenstein is to have realized just this point. He agreed with Candra that philosophical analysis of our actual concept of ourselves cannot yield a useful *revision* of the concept which can somehow free us from its bewildering implications. What Wittgenstein failed to understand, from Candra's point of view, is that what drives the philosopher's obsessive attempt to revise this concept, which encapsulates what Wittgenstein would call one of our most basic forms of life, is an underlying fear that we do not really exist! Nor does he seem to have conceived, any more than have most others in the Western tradition, the real possibility that our attachment to ourselves as we actually conceive ourselves brings about our suffering.

NOTES

[1] See Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984), p. 273 and Appendix J; Th. Stcherbatsky, *The Soul Theory of the Buddhists* (Delhi: Bharatiya Vidya Prakashan, 1976).

[2] The translation, to which I shall refer as the 'Refutation,' is by James Duerlinger, *Journal of Indian Philosophy* **17**: 137—187, 1989. Details about the text can be found in Duerlinger's introduction which precedes the translation in the same volume and is published under the title, 'Vasubandhu's "Refutation of the Theory of Selfhood,"' pp. 129—135.

[3] See Duerlinger's introduction, mentioned in previous note. A detailed comparison of Vasubandhu's version of the reductionist theory to the versions of Hume and Parfit will be undertaken elsewhere. Claus Oetke, in *'Ich' und das Ich* (Stuttgart: Franz Steiner Verlag Weisbaden GmbH, 1988), includes a paraphrase of Vasubandhu's chapter, along with an attempt to analyze it from a contemporary philosophical perspective. His treatment, however, is based on an understanding of theories of persons drawn from the work of Peter Strawson rather than from the work of Parfit, and it does not include, as mine does, an attempt to deal with Vasubandhu's theory as part of the Indian philosophical commentarial tradition.

[4] For our purposes we may adopt the traditional classifications of the Indian Buddhist philosophical schools as set out in *Tibetan grub mtha'* (*siddhānta*) texts. The mastery of this literature is a good first step in our attempts to relate Indian Buddhist philosophy to our own philosophical tradition.

[5] See Edward Conze, *Buddhist Thought in India*, 3rd ed. (Ann Arbor: University of Michigan Press, 1967), pp. 122—130.

[6] *Madhyamakāvatārabhāṣya*, Ch. 6, vv. 120—176, with commentary, as translated into Tibetan (*dbU ma la 'jug pa bsad pa*; P5263).

[7] I shall reserve for another occasion an account of Vasubandhu's intricate confrontation in the 'Refutation' with the nonreductionist theories of the Vātsīputrīyas, Sāṃkhyas, and Vaiśeṣikas.

[8] The apparent object of our actual concept of ourselves from the first person singular perspective is what, in most of Indian schools of thought, is called the *jīva*. Elsewhere I shall argue that it is skepticism concerning the real existence of a *jīva* that leads Indian philosophers to propound theories concerning the true nature of the self and its mode of existence.

[9] Revisionist theories may be straightforwardly revisionist or covertly revisionist. The latter incorporate a variety of arguments to the effect that such and such components of our actual concept of persons are not really components of that concept, but only seem to be so. To the extent that Western theories of persons can be assimilated to this Indian philosophical framework, in which the basic problem of self is set by skepticism concerning our real existence as actually conceived, most Western theories of persons would be classifiable as covertly revisionist. For Western philosophers tend simply to distinguish different meanings, senses, or uses of our first person singular concept of ourselves and ignore the fact that in our unphilosophical mode of consciousness we employ it to refer to ourselves in reliance upon our appearance to ourselves of existing independently with natures of our own.

[10] I shall argue elsewhere that Aristotle and Strawson would seem to hold realist versions of the nonrevisionist view of persons when judged from this perspective, and that Wittgenstein is best interpreted as holding a conventionalist version of the nonrevisionist view.

[11] *Reasons and Persons*, pp. 210—213.

[12] *Madhyamakāvatārabhāṣya*, Ch. 6, vv. 146—148, with commentary.

[13] *Reasons and Persons*, p. 241.

[14] There are other classifications which can be formed, but need not be for our purposes. Harold Noonan, in *Personal Identity* (London: Routledge, 1989), Ch. 5, in effect distinguishes, on the basis of the explicability of the relation which binds 'person-stages' into a continuum, versions of what I am calling reductionism and nonreductionism. His terminology differs from mine, in part, because his focus is entirely upon theories of personal identity. I shall consider below only the Buddhist statements of what I am calling the realist and modified realist versions of reductionism.

[15] What I have called here a modified realist version of reductionism might also be called a quasi-conventionalist version of reductionism, since the Svātantrikas claim, as Candra does, that all phenomena exist only by convention, except that in their case the conventional existence of a phenomenon does not exclude its possession of a nature of its own. The fullest account of the Svātantrika viewpoint in Western literature is to be found in Donald Lopez, Jr. *A Study of Svātantrika* (Ithaca: Snow Lion Publications, 1987).

[16] See 'Refutation,' p. 155.

[17] In the *Abhidharmakośa* Vasubandhu claims that all phenomena are either permanent or impermanent. But here "impermanent" is used both in the strict sense of existing only for a moment and in the loose sense of existing over time as a continuum of phenomena which are impermanent in the strict sense and are caused to retain the characteristics on the basis of which they are initially named and conceptualized. Consequently, he believes that no phenomena change in the sense in which change requires the existence of an unchanging substratum which undergoes a change of properties. It is only in the loose sense of "impermanence" that persons are included among impermanent phenomena. In my account of Vasubandhu's theory I am using the terms, "impermanence" and "impermanent," in their strict senses, and

when I speak of a causal continuum of a collection of phenomena on the basis of which a person is named and conceptualized I always mean a continuum in which phenomena produce others which preserve characteristics of the phenomena on the basis of which they are initially named and conceptualized.

[18] Parentheses are used here and elsewhere in the translation to mark off additions made by the translator to facilitate comprehension of the text.

[19] The term, "self" (*ātman*), is used here and in its next occurrence to refer to whatever is in fact a person. Later in the text Vasubandhu consistently uses the term in reference to persons conceived to possess an existence not reducible to the existence of the aggregates which comprise one's body and mental states. An extension of the normal use is that which occurs in "selfless" when used in the claim that the aggregates are selfless, for in this case the meaning is that none of the aggregates individually is either a self in the above sense or something possessed by a self in that sense. The term has its normal use here in the locutions, "seeing (oneself as) a self" and "clinging to (oneself as) a self."

[20] This is the basic logical framework of *Madhyamakāvatārabhāṣya*, Ch. 6, vv. 124–145, with commentary.

[21] *Madhyamakāvatārabhāṣya*, Ch. 6, v. 135cd, with commentary.

[22] *Ibid.*, Ch. 6, v. 127cd, with commentary. The Svātantrikas are thought to be committed to the idea that the aggregates really exist because they claim that they exist with a nature of their own.

[23] *Ibid.*, Ch. 6, v. 131, with commentary.

[24] The verbal form of this, *prajñapyate*, I render as "is named and conceptualized," instead of the more usual "is designated" and "is imputed." Claus Oetke, in '*Ich*' *und das Ich*, p. 196, fn. 96, renders the ". . . *prajñapyate*" construction as "*Es ensteht die . . . Konzeption*," and in so doing misses its ambivalence to the domains of discourse and conceptualization. A more accurate rendering, I believe, would be ". . . is apparently named and conceptualized," (and so also it would be more accurate to render *prajñapti* as "apparent name or concept,") but since this seems overly to complicate the translation, I have decided to use the simpler translation and caution the reader about inferring from it that there is in fact something named and conceptualized.

[25] For this general information and other general points concerning Indian Buddhist theories of persons, see, for instance, Geshe Sopa and Jeffrey Hopkins, *Practice and Theory of Tibetan Buddhism* (New York: Grove Press, Inc., 1976), Part II.

[26] David Hume, *Treatise of Human Nature* (Oxford: Clarendon Press, 1960), Bk. I, Part IV, Section IV.

[27] *Ibid.*, p. 252.

[28] I shall argue elsewhere that comparable problems plague the corresponding arguments of Hume and Parfit.

[29] See the 'Refutation,' p. 144.

[30] *Ibid.*, p. 163. He does not develop this criticism further, perhaps, because he thinks it sufficient to dismiss the Mādhyamika view on the ground that it contradicts the Buddha's denial of nihilism. He does not attempt to answer any of Nāgārjuna's criticisms of the view that phenomena really exist.

[31] My translations of "*saṃvṛtisatya*" as "a conventional deceptive reality" and "*paramārthasatya*" as "an ultimate reality," rather than as the more usual "conventional truth" and "ultimate truth," respectively, is an attempt to convey in modern

terms the etymologies the Indian Buddhist philosophers offer for the expressions. The sense I glean from the etymologies is that a reality is a phenomenon which possesses the mode of existence it appears to possess, a deceptive conventional reality is deceptive in that it does not possess the mode of existence it appears to possess, but by convention possesses the mode of existence it appears to possess, and an ultimate reality is a phenomenon which in fact possesses the mode of reality it appears to possess. A deceptive conventional reality, so understood, is no more a reality than a dead man is a man. Nonetheless, because its lack of reality is not recognized except by those who have developed the wisdom and concentration to perceive it directly, it is by convention called a reality, and this convention is followed by the Buddha.

[32] *Madhyamakāvatārabhāṣya*, Ch. 6, vv. 127–128, with commentary.

[33] *Mādhyamakakārikā*, Ch. 27, v. 16.

[34] *Madhyamakāvatārabhāṣya*, Ch. 6, v. 129, with commentary.

[35] *Ibid.*, vv. 134, 162, with commentary.

[36] *Ibid.*, v. 166, with commentary.

Philosophy Dept.
University of Iowa, U.S.A.

This page intentionally left blank

SELF-CONSTRUCTION IN BUDDHISM

David Bastow

I

In the section of the *Treatise of Human Nature* on 'Personal Identity' (Book I Part IV Section vi), Hume sketches a reductive theory of what it is to be a person. People 'are nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity.' Thus what is ordinarily thought to be a single unit is analysed into a multiplicity of smaller units.

The presentation of such a theory should contain three elements:

1. a specification of what actually exists, according to the theory;
2. an explanation of how the mistake in common understanding comes to be made, how a multiplicity comes to be thought of as a unity;
3. arguments showing that the theory is more likely to be true than its rivals.

1. What exist, according to Hume, are perceptions (the term used in a wide sense for any state of the conscious mind); these are constantly and rapidly changing, and succeeding each other. Those which are taken to constitute a single person are connected to each other by the relations of resemblance and causality. Memory provides a good example of resemblance. 'For what is the memory but a faculty, by which we raise up the images of past perceptions . . . An image necessarily resembles its object.' As for causation 'we may observe that the true idea of the human mind, is to consider it as a system of different perceptions or different existences, which are linked together by the relation of cause and effect, and naturally produce, destroy, influence and modify each other.' Hume does not mention it, but relations of resemblance and causality will also hold between perceptions ascribed to different people; but presumably the relations within one person's stream of perceptions are tighter.
2. How do we come to mistake what is really a multiplicity for a unity? The question concerns how we think about existence through time. Much of this Section of the *Treatise* is taken up

with an account of our general propensity to confuse identity through time with diversity through time. When we observe a succession of closely related objects, we tend to think of the succession as if it were a single object 'remaining invariable and uninterrupted'. The explanation is in terms of a psychological contingency: the action of the imagination in the two types of case is 'almost the same to the feeling'. 'The relation [between the successive perceptions] facilitates the transition of the mind from one object to another, and renders its passage as smooth as if it contemplated one continued object.'

How does this general doctrine apply to the particular case of personal identity? In this case, as already mentioned, the perceptions are closely related by resemblance and causality. In reviewing their succession, the mind passes smoothly along the sequence, fails to distinguish its own operation from the observation of a single uninterrupted object, and so invents such an object – the soul or self – to justify its mistaken belief in identity.

A comparison between paragraphs 1. and 2. above reveals an unexplained lack of fit. Paragraph 1. speaks of a sequence of extensionally related perceptions. Paragraph 2. has not only that sequence, but a mind passing, with deceptive smoothness, along the sequence; and a mind being deceived by this smoothness, coming to believe in an uninterrupted identity, and hypostatising an unobserved self. The question must arise, whether the primitive apparatus allowed by the theory can perform the sophisticated tasks required of it. How does the self concept get itself constructed?

3. Hume's argument for his theory – apart from his attempts to show that it is internally coherent, and that it explains what it sets out to explain – consists of the simple empiricist claim that there cannot be an identical uninterrupted self, to which all the perceptions belong, because no such self can be observed. All that can be observed is ceaselessly changing.

Several writers have remarked on the similarity between Hume's theory of personal identity and the Buddhist doctrine of 'no-self', with its explanation of the mistaken concept of self or person in terms of short-lived physical and psychological elements. Derek Parfit, in *Reasons and Persons*,[1] presents his own basically Humean theory of persons, and claims 'Buddha would have agreed'. I think

[1] Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984), especially Chapter 10.

that in the main Parfit is right about this; but Parfit's theory has the same lacuna as Hume's. I shall later describe Buddhist suggestions about how this may be filled.

Parfit's theory is more thorough-going than Hume's, in that the consequences of the breakdown of the concept of person are taken more seriously, and worked out in more detail. The elements of his reductionist theory are once again states of mind or mental events, related by what he calls 'psychological connectedness'. This notion is explained by means of examples; it holds between a seeming memory, and the experience of which it seems to be a memory; between an intention to do something and the doing of it; between having character traits, beliefs etc at one time and having the same or similar character traits or beliefs at another time. Parfit adds that these psychological relationships must, if the sequence is to come to be seen as a person, rest on some physical causal connexion; usually the existence of a continuing, though changing, physical body. These psychological and physical connexions are together called relation R. The relation of personal identity is the relation R plus the proviso that there is no more than one Person 2 who is R-related to Person 1, or Person 1 R-related to Person 2. But Parfit claims that we can easily imagine circumstances in which the proviso would not be satisfied; that is, people could fuse, or bifurcate in various ways; so we ought to concern ourselves with the basic relation R without the proviso.

Parfit argues for his theory in the Humean way: none of us is directly aware of himself as a 'persisting subject of experiences', a pure ego. But he also uses powerful new arguments, which consist of what may be called 'bizarre stories', about what might happen to people. They could be given different but similar bodies or brains, or different and dissimilar bodies or brains, or psychological characteristics with varying degrees of R-relatedness, or any combination of these, or they could in various ways fuse or bifurcate physically or mentally or both – and so on. The claim is that if, with reference to these stories, we ask 'Is it the same person?' we find that the concept of person breaks down under the strain; there are many cases in which the answer is not clear, or a decision either way seems arbitrary. That is, the concept of person suffers from what Parfit calls 'indeterminacy'. Parfit argues that indeterminacy entails a reductionist analysis. The stories are persuasive; and the indeterminacy which they demonstrate certainly shows that the concept of person and personal identity are complex; but it does not compel acceptance of Parfit's type of

reductivism. For Parfit, like Hume, fails to appreciate the deep implications of the obvious fact that we do look back on our experiences, and look forward to our intended actions, and see it all as a unity. Parfit says 'This unity does not need a deep explanation. It is simply a fact that several experiences can be *co-conscious*, or be the objects of a single state of awareness.'[2] He goes on to compare it to short term memory. But this extensionalist account is as inadequate for the self-concept as it is for memory. I shall argue, on well known Kantian lines, that a necessary part of basic psychological functions like memory and intentional action is the person's concept of himself; and this requires intensional relations, that is, relations which exist only insofar as they are understood to exist. In fact any theory of the person must involve an account of the person as an intensional agent; this is central because I agree with Kant and, as we shall see, with Buddhist accounts, not only that we as people must have self-concepts, but also that our self-conception has to be actively constructed. In fact we need to take into account what Kant would call active synthetic understanding.

It is perhaps worth recalling Kant's account of the conditions for memory. In fact he claims, more generally, that these conditions lie behind every case of conceptual understanding, for applying a concept always involves memory. Conceptualising always involves understanding that a multiplicity of intuitions are united by the fact that they follow a rule. Kant assumes that the initial awareness of the manifold must take some time; so at the time of understanding, data from various times must be 'run through and held together', an activity which involves memory. The past data must therefore be reproduced at the time of understanding – but reproduction (this is what Parfit calls 'co-consciousness') is not enough. The data must be known to be reproduced; they must be recognised. Kant claims, and I agree with him, that recognition must take a propositional form; something like '*I* experienced that in the past'. So consciousness of oneself as enduring through time is required. How is this possible? As Hume and Parfit say, such a self cannot be directly experienced as a kind of changeless crystal to which changing experiences can be severally attached; so an empirical self-concept must be constructed, under which the self can be thought of as to some extent stable, but to some extent changing, like other objects in the objective world.

Memory, therefore, needs self reference, and so some kind of self-

[2] Op. cit., 250.

concept; and this self-concept needs to be constructed. Construction is of course a synthetic activity; and it must involve the understanding. Whatever the elements or units are, they must be put together in a way which makes some kind of sense. For Kant, all known unity is understood unity. What forms then can this construction of an empirical self take? It must obey the categorical rules which makes possible any re-identification; but there are very many ways in which this requirement can be satisfied, as Parfit's bizarre stories show. We may think it wise, in the self-concept we construct, to allow for bifurcation, fusion, tele-transportation[3] etc; but what we cannot do without is *some* concept of ourselves. This is required for memory; and also, it seems obvious, for the link between intention and action, which gives or attempts to give an understood unity to the future, as memory does to the past.

Parfit does recognise that person-concepts exist;[4] but thinks about them in a purely objective way, as part of the furniture of the universe. He fails to realise that if persons can be the objects of (perhaps mistaken) thought and belief, they must also exist as active subjects, as doers, constructors, synthesisers. For Parfit, the idea of the person as subject is empty. This is because he represents the subject of experiences as that which 'has' them; if this is the subject's only role it is not difficult to argue, as he does with respect to the *cogito*, that 'someone has a thought' can be resolved into 'there is a thought'. But the subject as agent cannot be dismissed so readily; for things actually get done. In Kant's most abstract terms, manifolds get synthesised into unities, which thereby are understood as unities. This kind of agency cannot be reduced to causal relations between similar perceptions.

My argument so far has been negative. The Hume/Parfit theory of the person is inadequate because it does not allow for, or does not clearly allow for, the active construction of a self-concept. But if we can be sure that agents, synthesisers do exist, we still need to know what kind of existence they have. Do we have to reintroduce the Cartesian ego as subject? This seems ruled out by the arguments for indeterminacy. If we turn to Kant, the ontological status of what for him is the subject of synthesis, that is presumably the original unity of apperception, is far from clear. Could there be a theory more sophisticated than Parfit's, allowing that self-concepts are constructed, and so a person as constructor is needed, but also

[3] Op. cit., 199–201.
[4] Op. cit., 210.

allowing that the construction may be variously done, so that there may be a variety of answers to 'Is this the same person as that?'? Is it possible to explain the relationship between person as constructor, and person as constructed?

II

As one possible answer to these questions I should like to present a rather free and speculative interpretation of the theory of the person put forward in the Pali Buddhist texts. Parfit claims that this theory is similar to his own, and there is certainly some justification for the claim; but the Buddhist account is, at least potentially, rather more sophisticated than he suggests.

We may start with the earliest of the Buddhist texts which Parfit himself quotes.[5] The verses are spoken by a Buddhist nun to Māra the tempter, who asks 'Where do *sattas* (literally 'beings': people, gods, demons etc) come from? How do they arise and pass away?'

> Why do you keep referring to 'beings', Māra?
> You have gone in for a false view of things.
> There is merely a pile of *saṃkharās* here;
> No 'being' is to be found here.
> Just as the bringing together of parts
> Gives rise to the word 'chariot',
> So with reference to the *khandhas* that are present
> Arises the conventional term 'being'.[6]

Obviously the view being put forward here is reductivist in a strong sense; the analysis of 'beings' into *saṃkhāras* or *khandhas* is taken to imply the ultimate non-existence of 'beings'. Speaking of them is using what is no more than misguided convention. This doctrine is usually known as *anattā*, no-self; each of us must apply it to himself.

I mentioned at the beginning of this paper three requirements we might make of a reductive theory of the person: 1. an account of what exists, according to the theory; 2. an explanation of how common understanding comes to be mistakenly non-reductive; and 3. arguments for the truth of the theory. We saw that Hume and Parfit failed to explain adequately the relationship between 1. and

[5] Op. cit., 502.

[6] *Samyutta Nikāya* (London: Pali Text Society, 1884), volume 1, 135. The translation is mine, as are the other quotations from Pali texts, except where other translations are mentioned in these notes.

2. A theory which hopes to do better will have to produce, for 1, an account of the elements of the reduction which is rich enough to fulfil the task set in 2. The previous paragraph gives a preliminary answer to 1.; in the rest of the paper I shall expand and explain this answer to show how it can cope with 2.

I shall not attempt to deal adequately with Buddhist answers to 3. They are a combination of soteriological, experiential and intellectual considerations. At the abstract intellectual level, it is interesting to note the emphasis placed on seemingly empiricist arguments. The doctrine of *anattā* is closely linked to the claim that all that we can see to exist in the world, including any aspect of ourselves by which we may try to identify ourselves, is impermanent, short-lived. The *khandhas* are a five-fold classification of the elements of which 'beings' are made up: physical form, feelings, perceptions, motives and consciousness. The argument is repeatedly used, of these five, 'Physical form (and the other four) are impermanent; what is impermanent cannot constitute a self'. The possibility of the existence of anything else, outside the *khandhas*, which is permanent and therefore could be a genuine object of self-reference, is discounted by saying that any such thing is inexperienceable and inconceivable.

But to return to our main concern; a theory which analyses a long-lived person into short-lived constituents must contain an explanation of how the constituents are temporally related to each other – how the appearance of the abiding self comes about. There are hints of this in the potent word *saṃkhāra*.[7] The term is used in the verses as a general indication of the nature of the elements. Its components are the prefix *sam* meaning 'together', and *khāra* from the root *kṛ*, to do or to make; its uses are always related to synthetic activity. In this context it indicates that the elements are those things which are heaped together to constitute the person; but also that they are not themselves passive in this process; they are putters together as well as being put together. Certainly there is no external agent to bring about the synthesis.

How they are able to do this is not made clear in the earliest Buddhist texts. The theory of relations in these texts is called what may be translated 'conditioned co-origination' (*paticcasamuppāda*), that is, events arise conditionally on others; but the nature of the conditionality, extensional or intensional, is not explicitly discussed.

[7] There is an extended and profound consideration of this concept, and its place in the history of Indian thought, in Lilian Silburn, *Instant et Cause* (Paris: Librairie Philosophique J Vrin, 1955). See especially Chapitre VI: Bouddhisme Canonique.

The standard texts describe how the sequence of origination (always associated with suffering; always, for the Buddhist, to be brought to an end) prolongs itself. Because of ignorance of the true nature of things, sensory experience, contact of consciousness with the outside world, gives rise to pleasant or unpleasant feelings, which themselves bring about craving, or thirst, that the pleasant feelings may be prolonged, the unpleasant brought to an end. It is in the arising of craving that subject and object come to define themselves in opposition to each other; and this opposition, entailing self-consciousness, is projected into the future, as craving becomes a dependence which is habitual and grasping; occurrent desire becomes a long-term project of appropriation. Thence arises the stream of life, extended indefinitely, on the Buddhist view, by rebirth. This process is sometimes called *ahaṃ-kāra*, I-making, the putting together of what thinks of itself, and is concerned with itself, as a continuing person. But knowing and seeing that none of the five *khandhas* is the self, that they are indeed impermanent and painful, 'there come to be no latent tendencies to I-making, to mine-making, with respect to this consciousness-informed body and all the phenomena external to it'.[8]

What we can take from these passages is that the notion of an I, a self, is one that is constructed, put together; the process of I-making is comprehensible, regular, takes place according to conditions; but it is not taken to be inevitable. In fact the Buddha thought that his discovery of the conditions gave him the ability to avoid their results by removing the conditions. 'This being, that becomes: this not being, that does not become.' But we are still far from understanding the causality of this construction, or indeed the true nature of the elements of the construction. The self projection, and therefore self conception, at the end of the process, somehow arises out of the short-lived non-personal constituents of sensory contact and affective reaction; but to assert that this process of self-construction is conditioned by ignorance is not to answer the theoretical question of how the 'I' is constructed. Who, we still want to ask, does the construction?

What is needed, and what Buddhism provided, especially in its later more scholastic writings, is a theory of the elements out of which the construction is made. Obviously no separate constructor can be invoked – we do not have a housebuilder as well as the bricks; so the elements must themselves contain the seed of the

[8] *Majjhima Nikāya* (London: Pali Text Society, 1977), volume iii, 18–20.

construction. The elements are all there are, impermanent and unsatisfactory as they may be.

We may remember here the central Kantian point that construction is an intensional activity – it involves the bringing about of a unity in the sense of something understood to be a unity. The construction of a self therefore cannot but involve the synthesis of a concept of the self; though it will also involve the putting together of elements which fit in with the rules of this self-concept, and perhaps the creation of new elements especially designed to fit into it.

It is not enough, then, to characterise the elements as dynamic, causally active; the causality must be of a particular kind. The elements must have some of the features of an active mind. But also we must be able to account for the Buddha's emphatic assertion that they are in themselves void of a self.

I shall first give an abstract description of what could satisfy these conditions; and then elaborate it according to Buddhist models. We could call the basic type of element a state of consciousness, but this would have implications which are too static and passive; so I prefer to use the more dynamic phrase 'rise of thought' (*cittuppāda*), which is the most abstract mental concept in the major scholastic work called *The Enumeration of Basic Entities* (*Dhammasaṅganī*).[9] A thought arises at a point in time (and space, since it is always linked to a body of some kind). But because it is a thought, and therefore intensional, it has a content. In fact it will take the form of a synthesis, a proposition; which itself has a content, namely, the manifold which is synthesised. What is to be synthesised may and usually will stretch beyond the time and place of the arising thought itself; it may include thoughts and physical events from the past and the future. Thus things happening at several times may be linked together, seen to constitute a unity.

In fact the Buddhist texts describe, with varying explicitness, three basic possibilities for the arising thought. I list the three here, and then discuss them in more detail. The first is that with which we ordinary people are, without realizing it, most familiar; the arising thought can go in for person-making, and thus become tied into a sequence of thoughts and experiences, all based on desire and all resulting in suffering. The second is manifested, in the Buddhist path, in the state of mystical trance which is sometimes

[9] *Dhammasaṅganī* (London: Pali Text Society, 1978), 1. The concept of *cittuppāda* is discussed in Alexander Piatigorsky, *The Buddhist Philosophy of Thought* (London and Dublin: Curzon Press, 1984), Chapter 4.

called *samādhi*; here the arising thought abandons all temporal outreach, and concentrates on itself alone. This may, as we shall see, imply an inwardly focussed synthetic activity; or the cessation of all synthesis. The third possibility is of a synthetic understanding unmotivated by desire, totally dispassionate. To outward appearances this produces a person, but one completely detached from the world in which he lives. This is the saint or *arahant*, or the Buddha himself – though such personifying modes of reference are not really appropriate. 'They live in emptiness, in the unmarked, in freedom; difficult it is to trace their course, like that of birds in the air'.[10] Sometimes the texts imply that the third possibility, arahantship, is the goal, and the inward-turning synthesis of *samādhi* merely a means to it. This is the interpretation I shall develop; though the texts do also contain the more nihilistic suggestion that the cessation of all synthetic activity is the true goal.

The first possibility then is self-construction, I-making; that is, the making of a being which thinks of itself as a self. The Buddhist view is that a special kind of synthesis goes to produce such a self-concept, namely one based on or motivated by desire or craving (the Pali term is *taṇhā*, literally meaning 'thirst'). The understanding of the temporally extended manifold as a unity in this case is accompanied by an affective relationship between the arising thought and the idea it constructs – that is, there is self-concern. As we shall see, this link between synthesis and desire is, though usual, not necessary. I am not able to give an adequate abstract characterisation of desire; in this scheme of things it must be logically prior to the desiring person; as must ignorance, with which it is inextricably associated. So the term is shorn of its usual logical associations, but is not provided with a new set; unfortunately its status in my account is that of a basic undefined term. Examples of its manifestations, to be given later, will have to serve to explain it. Certain special kinds of direct relationship with past and future, namely memory and anticipation, are particularly important in self-making; but do not compel it. That is, memory and anticipation *may* (in terms of the third possibility) be in a sense non-personal, or trans-personal, containing no desire in their make-up, merely what we may call dispassionate understanding. So, though the arising thoughts are themselves non-personal, or

[10] *Dhammapada* (London: Pali Text Society, 1914), stanza 92. For a generally reliable translation, see *The Dhammapada*, translated with notes by Ṇārada Thera (London: John Murray, 1972).

pre-personal, persons may be constructed by and out of them, by 'running through and holding together', in a desire-laden way, a sequence of past, present and future thoughts and physical events. This is like other strongly reductive analyses, in that its evaluative effect is to downgrade the analyzandum, the concept of person; but in Buddhist thought the fact that the constituents are non-personal is no reason for disparaging *them*. They are potential person-constructs, and so potentially enmeshed in suffering; but equally they are potential person-transcenders, capable of liberation.

The term 'I-making' covers two kinds of making: making something of a manifold, like making sense of an arrangement of coloured pigments at an art gallery; but also making something in the way that one makes a chair, shaping the component parts so that they can be fitted together according to a plan. That is, the thought, considered in itself, is potentially active in two ways; the first, already described, is the unifying activity of the understanding. But construction of a personal unity usually involves not just seeing an already existing manifold as a unity, but organising the future so that it will fit into the individual's self-conception. So desire moves to action, perhaps action in the long term, as plans are made, commitments undertaken. Action in the world obviously involves, though does not consist entirely of, physical causality – or at least some kind of causality which is extensional; its laws are independent of people's understanding of the matter. It is usually[11] assumed that the causality linking what in misleading personal terms is called rebirth is a kind of law of nature; that is, the causal links are there whether anyone knows about them or not. But if rebirth is a causal consequence, it has as a necessary condition the intensional activity of I-making. So putting an end to I-making implies putting an end to rebirth, to the circle of *saṃsāra*. If there is no desire, and therefore no I-making, there is no motivation to action, so the sequence of conditioned co-origination does not take place.

Self-making is a grasping synthesis of past and future, based on desire. The texts usually link desire with ignorance, and we remember that the first element in the standard list of conditioned co-origination is ignorance. It seems that the Buddha had, and expected that his followers would be able to have, a vision of things 'as they really are', free of mistaken speculative theorising, both that which went on in rival philosophical traditions, and that which he thought the ordinary man was guilty of, in his self-construction.

[11] Though not always; see Piatigorsky, op. cit., 76–81.

How then are things really, on the account we have been describing? Do people really exist, or do they not? Should we say, in the reductive mode, that arisings of thought exist, but people do not – or only exist as a façon de parler? Does the house exist, or should we say that only the bricks really exist? The dichotomy 'exists/does not exist', though seemingly exhaustive, is too crude to say what should be said here. Thoughts may but need not construct a person – this is how things are. The vision of reality is of persons with their mode of construction laid bare. The Buddha said that he did not assert a theory of existence; nor of non-existence; nor of unity, nor of plurality: 'not approaching any of these extremes, he teaches a doctrine by a middle way';[12] the doctrine is the conditioned co-origination account of self-making.

The ignorant man though does not understand how he himself is constructed, nor that his construction is only one of the possibilities for the thoughts which he calls his, nor that suffering is the inevitable result of this desire-ridden construction. So attachment to past and future comes about. This embraces a general concern for oneself in past and future. The person with a misdirected mind thinks 'Was I, or was I not, in some past period? What was I, how was I, what did I become? Will I or will I not come to be in a future period? What will I be, how will I be?' etc.[13] Attachment to past and future involves desires motivating action. 'Someone considers and turns over in his mind thoughts about things brought about by desire and lust in the past. As he does so, desire is born, and so he is fettered by those things. This I call a fetter, this thought full of desire.'[14] The contrasting teaching is 'Let go the past, let go the future, let go the present: crossing to the future shore of existence, with mind released from everything, do not again undergo birth and decay'.[15] This passage refers to the end of rebirth, but it will be noted that this is not something the traveller on the Buddhist path, the path leading to the extinction of desire and therefore of self-making, can properly wish for – that he, he himself, should come to an end of rebirth, or even successfully complete the path.

Does this condemnation of involvement with past and future extend to moral action, i.e. does concern with the past and future interests of other people equally manifest craving and ignorance? We must distinguish between presentations of the ideal, descriptions

[12] *Majjhīma Nikāya*, volume ii, 76–77.
[13] *Majjhīma Nikāya*, volume i, 8.
[14] *Anguttara Nikāya* (London: Pali Text Society, 1964), volume i, 264.
[15] *Dhammapada*, stanza 348.

of the Buddha's vision and its implications, on the one hand; and prescriptions for those who while sharing the Buddhist ideals are still locked in personhood, on the other. From the standpoint of the ideal, all reference to personal life, and so all concern with the relationship between 'oneself' and 'other persons' is based on ignorance, and is to be condemned. But that does not mean that no discriminations can be made between styles of personal life. On the path towards the ideal, those qualities are encouraged which point to the ideal. Selflessness in the moral sense is not unrelated to the metaphysical doctrine of no-self; it obviously connects with the third rather than the second of my three possibilities, that is with desireless synthesis, rather than with the thought's non-temporal self-absorption. There are meditational disciplines for people on the path which systematically extend the emotions and attitudes of loving-kindness, compassion, sympathetic joy and even-mindedness from those for whom one naturally has such feelings to the whole world without discrimination.[16] This is formally similar to Parfit's suggested move from M, a morality which gives a special place to one's kith and kin etc, to N, a neutralist morality which has no place for such extended egocentricity.[17] On the path, then, though all concern for past and future is condemned, what we would call moral concern for what has happened and will happen is preferable to purely selfish concern.

The second of the possibilities is that the synthesised content of the arising thought does not extend outwards in time, calling in elements from past and future. The thought therefore can be concerned only with itself. 'The noble one is he for whom past, future and present are as nothing; he has nothing and grasps after nothing.'[18] The temporal reference even of the present disappears, when there are no other times to contrast it with.

In this context Buddhist conceptions of mental training are very relevant. There is frequent emphasis in the texts on control; and the metaphor of taming – discipline to bring about control – is often invoked. The implication is that the natural tendency is for thought to run after the past and future; it has constantly to be narrowed down to the present. This is obviously central to the technique of 'mindfulness' (*sati*) – concentrating on some short term and repeated event like one's breathing, vigilantly noting and dis-

[16] Buddhaghosa, *The Path of Purification* (translation, by Nyānamoli, of *Visuddhimagga*) (Berkeley and London: Shambala, 1976), chapter XI, §9.

[17] *Reasons and Persons*, 108–109.

[18] *Dhammapada*, stanza 421.

passionately setting aside the wider-ranging thoughts which intrude. Those who have tried this technique will readily appreciate that the high value placed on the calm of the pure concentrated mind can be seen as metaphysically significant; allowing one's thinking to chase after other times and other places comes to seem an encouragement of what are called 'adventitious defilements'. The only true reality is here and now. In the texts there seem to be two attitudes and developments of this inward-turning possibility. The first is to exploit the absence of contrast, of any kind of duality, to dissolve away even the barest synthetic activity, in order to reach what from a worldly point of view can only be called extinction. Thus there is a progressive technique of concentration (the *jhānas*) which culminates in meditation on boundless space, then boundless consciousness, then nothingness, then neither perception nor non-perception. But in general the *jhānas* are treated as means to a further end, and not even a necessary part of the path. The more positive treatment of the thought which does not go elsewhere for its content is that it is an intense alert vision of how things are; the thought sees itself to be impermanent, devoid of a self; and the significance of this has to be extended to what in personal terms would be called the whole of life.

The extension of this vision to constitute a way of living is the third possibility, which we must now consider. This is that a life can, and ideally should, be lived without person-making; more formally, that synthetic understanding and even some forms of activity in the world could be without desire, and so non-personal. In the Buddhist context, it would be claimed that ideals related to the value of personal life have here been transcended. It must be admitted that the Pali texts give us – that is, give people who have not even started on the path – only indications of what this possibility amounts to. There are two main aspects – what may be called the moral and the cognitive. In the Mahāyāna tradition of the *bodhisattva*, there is an explicit and much elaborated doctrine of the identity of these two aspects of the ideal, that is of the absolutes of wisdom or insight, and compassion; but descriptions of the ideal are not much elaborated in the Pali texts, which say much more about the path, from the point of view of those still engaged in it, than about the goal.

On the moral side, two types of being represent the ideal; the *arahant* or saint who has successfully followed the path to its end, and the Buddha, who himself discovered the path and then taught it to others. What kind of life do these non-people lead? With

respect to the Buddha, although in formal terms, as he said himself, nowhere does he exist, in fact the texts give us a kind of biography, according to which he lived a life of compassion, travelling and preaching, showing himself sensitive to the individual needs and circumstances of those he spoke to and lived with. That is he can, from our point of view, be seen as a moral hero, a manifestation of the relation I suggested earlier between moral and metaphysical selflessness.

For a more abstract description of what I have called the moral ideal we may turn to some verses from the *Sutta Nipāta*.

> 'Having what vision and what virtuous conduct is one called "calmed"? Tell me this, Gotama, when you are asked about the supreme man.'
> 'With craving departed before the dissolution of the body' said the Blessed One, 'not dependent upon the past, not to be reckoned in the present, for him there is nothing preferred in the future. Without anger, without trembling, not boasting, without remorse, speaking in moderation, not arrogant, he indeed is a sage of restrained speech.
> . . . It is not because of love of gain that he trains himself, nor is he angry at the lack of gain. He is not opposed to thirst, nor is he greedy for favour, being indifferent, always mindful. He does not think of himself as equal in the world. He is not superior nor inferior. For whom there exists no craving for existence or non-existence, him, indifferent to sensual pleasures, I call "calmed". In him there are no ties; he has crossed beyond attachment.'[19]

These mainly negative characterisations do suggest a style of life, one of complete self-control and self-reliance. 'Friendly amongst the hostile, peaceful amongst the violent, unattached amongst the attached'.[20] As already quoted, his path is as untraceable as that of birds in the air. He is in fact, as we should expect, a transcender of the categories of ordinary life.

In a few passages, the question of whether *arahant* and Buddha go in for action is addressed. Superficially, it is obvious that they do, otherwise continued life in the world would be impossible for them. But the concept of action and the associated concepts of planning, willing, intending, are so readily associated with the ideas of

[19] *Suttanipāta* (London: Pali Text Society, 1965), stanzas 848–857. Translated by K R Norman in *The Group of Discourses* (London: Pali Text Society, 1984), 142.
[20] *Dhammapada*, stanza 406.

grasping, attachment, desire, that a positive doctrine of action without desire, though surely needed, is not forthcoming.

The usual account is of the contrast between the ignorant man who craves and therefore acts, and the wise man who desires nothing and so grasps after nothing. 'He does not plan an act of merit or demerit. Not planning, not willing, he grasps at nothing whatever in the world. Not grasping he is unperturbed; he is of and by himself utterly well'.[21] Two models which might be developed to fill the gap are those of spontaneous action, and rational action, again thinking of Parfit's attempts to link rationality and neutralist morality, but I shall not attempt such development here.

The cognitive ideal, as present in the Buddha and to some extent in the *arahant*, is synthetic understanding of time, past present and future, free of any personal reference. The Buddha claims to be able to see the rebirth histories of all men – presumably by a kind of quasi-memory. He also over-comes the distinction between self-knowledge and knowledge of other minds; he can see into the hearts of all men. At another level, he has seemingly direct knowledge of the laws governing phenomena of *saṃsāra* – he knows that all *saṃkhāras* are impermanent and devoid of self, but he also knows that they are linked together in conditioned co-origination.

Understanding which is thus distanced from any personal point of view has obvious affinities with scientific understanding; and one thinks again of the Western model of rationality. This is often seen, both in practical and in epistemological contexts, as a transcendence of the limitations of the point of view of the individual self.

III

These then are the three possibilities for the arising thought presented in the Pali Buddhist Tradition. I have described them for what seems to me their intrinsic interest, though I can see that the claim on which they rest, that person-construction depends on desire, has not been adequately explained.

But the basic ontology, of an arising thought going in for synthesis, does not depend on this claim. Provided the basic Kantian conditions for synthesis are satisfied, the Buddhist account of the arising thought makes possible a wide variety of views about what one should think oneself to be, and therefore about criteria for

[21] *Majjhima Nikāya*, volume ii, 82.

re-identification of persons. That is, there are many different rules under which the arising thought can unite past and future events and thoughts, and see them as a self. The Buddhist view condemns all self-constructions, as entailing unsatisfied desire and therefore suffering; but we can now consider a theory detached from this extreme position, a theory of self-construction which allows for the indeterminacy with respect to 'person' which Parfit's bizarre stories argue for. Questions like 'am I the same person as David Bastow 40 years ago?' or 'does one stay the same person if tele-transported?' cannot be answered merely by reference to the basic ontology. Our culture has constructed provisional answers to them, and may in the future construct different ones; we as individuals may, for our own purposes, prefer yet other self-conceptions. With reference to the problems raised at the beginning of the paper, postulating the arising thought as the basic element of the theory makes it clear, as the Hume/Parfit analogies do not, how self-concepts can be constructed.

It may be remarked, finally, that the active role which individuals plays in constructing their own self-conceptions has been emphasised by existentialists. Kierkegaard, for instance, thought that unless one actively took responsibility for one's past, and even for one's heredity and environment, and projected one's plans for oneself into a limitless future, no real self had been created. Sartre emphasised, in a way which echoes Buddhist talk about the transience of all *saṃkhāras*, that decisions about oneself have constantly to be re-made, re-affirmed. As each new thought arises, the past becomes mere facticity, to which the for-itself remains committed only if it chooses to be. Each new arising thought provides a wholly new opportunity for self-construction.

University of Dundee
Dundee DD1 4HN
Scotland

BUDDHIST REDUCTIONISM

Mark Siderits

Department of Philosophy, Illinois State University at Normal

There has been much recent discussion concerning the relation between Parfit's Reductionism and the view of persons to be found in early Buddhism and Abhidharma. Some have claimed that Parfit is wrong to see in the Buddhist position an important anticipation of his own view—since the Buddhist view is not, as he supposes, Reductionist, but rather Eliminativist.[1] Others suggest that while certain of the Abhidharma schools might have held a Reductionist position, early Buddhism may defensibly be interpreted as Non-Reductionist in character.[2] It is something of a cliché among scholars of Buddhist and comparative philosophy that contemporary philosophical discussions would be enriched through greater familiarity with the Buddhist philosophical tradition. If such claims are to be taken seriously, then we should at least be clear about how the Buddhist philosophical problematic maps onto contemporary discussions. Further, we need to demonstrate, through concrete case studies, that Buddhist resources can facilitate genuine progress on current disputes. I shall here propose a taxonomy of views on persons and personal identity that I believe helps clarify the relations among the various Buddhist and contemporary positions on these issues. I shall also seek to show that when we develop a taxonomy that accommodates both traditions, certain of the currently held positions come to appear more plausible.

We might do well to begin with a general characterization of reductionism.[3] This is widely held to be, first and foremost, a view about what belongs in our ontology. To be a reductionist about things of kind K is, on this view, to hold that while it is not wholly false to claim that there are Ks, the existence of Ks just consists in the existence of certain other sorts of things, things that can be described without asserting or presupposing that Ks exist. Thus a reductionist about mobs would maintain that while mobs may be said to exist in a sense, the existence of a mob is really nothing over and above the existence of certain particular persons behaving in certain ways at a certain place and time. The characteristic reductionist "just consists in" clause is often explicated in terms of the "complete description" test: that we could give a complete description of all the facts about reality without ever mentioning Ks. Thus, suppose that last night a mob set fire to government buildings in the capital. Our reductionist about mobs would claim that we can exhaustively describe this event just by describing certain particular facts about the particular people in that mob, and without ever mentioning the mob. Since the mob's action of setting fire to those buildings just consists in certain individual actions, nothing is omitted from our description when we describe those individual actions and say nothing of a mob engaging in

arson.[4] And, claims the reductionist about mobs, if we can give a complete description of reality without either asserting or implying that mobs exist, mobs are thereby shown to be ontologically superfluous, and thus have no place in our ultimate ontology.

When reductionism is seen in this way, non-reductionism is then readily characterized as the view that things of kind K do belong in our ultimate ontology—that there is nothing that Ks "just consist in," since Ks are ontologically primitive. (Hence a non-circular analysis of "K" cannot be given.) But the third possible view about Ks, the eliminativist view, is not so easily captured on a purely ontological approach to the issue. For then the eliminativist about mobs would seem to be one who proposes that we eliminate mobs from our ontology altogether, and it is far from clear how this position would differ from that of the reductionist about mobs. It is only when we bring in the semantic dimension of the issue that these two types of positions can be clearly distinguished. To say, for instance, that organic chemistry may be reduced to quantum mechanics is to make a certain claim about the relation between two theories: talk of covalent bonds may be systematically replaced by talk of certain quantum mechanical states, so that the latter theory is thought to explain the predictive success of the former. By contrast, the demonic-possession theory of disease does not reduce to the germ theory, for there is no way systematically to translate talk of being possessed by a certain demon into talk of bacterial or viral infection.[5]

Thus it is that we take a reductionist stance toward the covalent bond, but an eliminativist stance toward disease-causing demons. The term "covalent bond" is now revealed to refer to certain distinctive sorts of quantum mechanical phenomena. So while a complete description of reality need not mention covalent bonds (whereas it would have to mention such things as quantum shifts), we may tolerate talk of such things just because the term is a useful way to refer to a certain class of quantum phenomena in which we take an interest. With demons, though, things are quite different. When we come to accept the germ theory of disease, it becomes apparent that our former talk of being possessed by demons cannot be seen as just a rough-and-ready way of referring to bacterial or viral infection. For one thing, demons are thought to have properties, such as malicious intent, that could not be explained on the assumption that they just consist in masses of microscopic replicators. For another, a given kind of demon was thought to be responsible for what we now consider to be essentially unrelated diseases (the viral versus bacterial pneumonia problem). Whereas the covalent bond is a posit of a useful though subsumed theory, the demon is a posit of a discredited theory; hence all talk of demons is to be eliminated.

Distinguishing between reductionism and eliminativism requires

introduction of the semantic dimension. But this complicates matters significantly. For success at translating between one theory and another is something that admits of degrees. Consider the terms "sunrise" and "sunset," which are intermediate between the case of the covalent bond and that of the disease-causing demon. We might have thought, when we transferred allegiance from the geocentric model to the heliocentric model of the solar system, that these terms were ripe for elimination. Yet they survive. Had we expected otherwise, this would have derived from the notion that their meanings seemed inextricably bound up with the now discredited geocentric theory. Instead, these terms exhibited sufficient semantic flexibility that we could retain them while suppressing the implication that the astronomical phenomena are explained by the sun's motion around the earth. This semantic shift was not accompanied by a shift in supposed referent: we take the Ptolemaic astronomers to have been referring to the same thing we refer to with these terms. We can imagine circumstances under which something similar might have occurred with our talk of demons (and as did happen with "humor"). What this suggests is that reductionism and eliminativism represent the ends of a continuum, with a middle range of cases in which it may be indeterminate whether the entities of the old theory are being reduced to, or eliminated in favor of, the entities of the new theory. But as is often the case with sorites phenomena, the existence of such an intermediate gray area need not count against there being a real distinction to be drawn between reductionism and eliminativism.

Characterized semantically, then, non-reductionism about Ks will be the claim that Ks will be mentioned in our final theory about the ultimate nature of reality. Both reductionists and eliminativists deny this claim, but they disagree over whether continued talk of Ks will have any utility in the light of our final theory. The eliminativist, of course, proposes eliminating all talk of Ks, both in our final theory and in its ordinary-language adjuncts. It might be thought, then, that the reductionist sees matters thus: while the term "K" is in principle eliminable from our language (since we can give a complete description of reality without mentioning Ks), its continued use is both tolerable (because of translatability), and of some utility given our interests.

Now I believe this is an accurate portrayal of the reductionist position, but there are those who would dispute this. Some draw a distinction between strong and weak reductionism. The strong reductionist (or "conventionalist") is said to subscribe to the eliminability thesis, while the weak reductionist does not. A weak reductionist about Ks maintains that Ks have a kind of "conceptual priority" such that, although a noncircular, nontrivial analysis of the term "K" can be given, some of the concepts occurring in the analysis will be adjectival on "K," in the way in which "citizen" is adjectival on "nation."[6] Thus the weak reductionist

denies what the strong reductionist affirms: that Ks are ontologically parasitic on, or logically constructed out of, more particular things; that we believe Ks to exist only because of the way in which we talk (hence the label "conventionalist" for strong reductionism). I have already indicated that I find this distinction problematic. It is not clear to me that so-called weak reductionism is properly characterized as a kind of reductionism at all. But further discussion of this point must await an examination of the reductionist, non-reductionist, and eliminativist approaches to the concepts of a person and personal identity, to which we now turn.

Parfit's general account of Reductionism (1984, pp. 210–214) provides a convenient starting point for this examination. Parfit characterizes Reductionism about persons as the view that the existence of a person just consists in the existence of a brain and body, and the occurrence of a series of interrelated physical and psychological events. Given the reductionist force of this "just consists in," all of the facts to which the existence of a person is here being reduced—the existence of a particular brain, the occurrence of a certain psychological event, the holding of causal relations between a particular psychological event and certain other physical and psychological events, and so forth—will allow a completely impersonal description, that is, a description that neither asserts nor presupposes that this person exists. Hence the Reductionist view of the identity over time of persons: that this just consists in more particular facts, facts that can be described in a thoroughly impersonal way. (This general schematism is fleshed out by neo-Lockeans with purely psychological facts, by others with facts about the body or the brain; hence there are two possible Reductionist approaches to personal identity.)

There are also two possible Reductionist views about what a person is: that a person just is a particular brain and body and a series of interrelated events, and that a person is an entity that is distinct from a brain and body and such a series of events. To those familiar with the Buddhist problematic, this will sound rather like the question whether a person is identical with or distinct from the five skandhas (bodily constituents, feelings, perceptions, volitions, and consciousnesses). And since the standard Buddhist response to this question is that the person is neither, one might be inclined to wonder how much affinity there is between Reductionism and any Buddhist view. Our puzzlement is heightened when Parfit also tells us that on the "distinct entity" view that he prefers, not only may persons be said to exist, they may also be said to be the subjects of experiences, the owners of their bodies, and the like. And yet all Reductionists, we are told, deny the Non-Reductionist thesis that a person is a separately existing entity, that is, an entity whose existence is distinct from that of a brain and body and a series of interrelated events.

Is Parfit's preferred form of Reductionism coherent? Indeed it is. The key to understanding the view may be found in this passage:

> Even Reductionists do not deny that people exist. And, on our concept of a person, people are not thoughts and acts. They are thinkers and agents. I am not a series of experiences, but the person who *has* these experiences. A Reductionist can admit that, in this sense, a person is *what has* experiences, or the *subject of experiences.* This is true because of the way in which we talk. (1984, p. 223)

The implication, of course, is that we conceive of persons as existing distinct from their experiences, and so forth, *only* because of the grammar of our language. A person is not in fact an entity that exists separate from a particular body, and so forth; our belief that they are is merely an artifact of our language, yet it is not entirely wrong to continue to speak of persons in this way. What is at work here is the tension characteristic of all reductionisms between a deflationary tendency that seeks to diminish the significance accorded the things being reduced, and a felt need not to sever all connections with existing practices. Since persons are not to be found in our ultimate ontology, the continued existence of a person cannot have the kind of importance it is commonly thought to have. Yet since our ordinary ways of conceiving of persons are not wholly mistaken, we may expect that at least some of the normative weight that we now invest in persons will be preserved in the more particular facts to which the existence of persons is to be reduced. Reductionism is a middle path between the extremes of Non-Reductionism and Eliminativism.

A Non-Reductionist need not maintain that persons are separately existing entities. Parfit distinguishes between two versions of Non-Reductionism: the view that persons are separately existing entities (such as Cartesian Egos) and the view that while we are not separately existing entities, the existence of a person involves a further fact, over and above the "more particular facts" of the existence of a brain and body and the occurrence of a series of interrelated physical and psychological events. The classic expression of the first version in the Western tradition is, of course, Descartes' conception of the "I" as a thinking substance; Swinburne (1984) represents a more recent formulation. (This type of view is also well represented in the Indian tradition, for example in the Nyāya theory of the ātman and the Sāṃkhya theory of puruṣa.) The basic idea here is that the human being (and possibly other life-forms as well) is a complex system consisting not only of those things, such as body parts and mental events, that are ordinarily observable through sense perception and introspection, but also containing some one entity that constitutes the essence of the system, that one part the presence of which is required in order for the system to exhibit the properties that we ascribe

to persons; it is the continued existence of this entity that constitutes personal identity over time.

While this type of view is relatively familiar, the "further fact" version of Non-Reductionism seems rather more puzzling. Here we appear to have a reluctance like that of the Reductionist to posit extra entities, together with the insistence that certain key facts about persons cannot be accounted for in terms of more particular, wholly impersonal facts. In the Western tradition, Reid, Butler, and more recently Chisholm have put forward views that might be interpreted as of this sort. But in each of these cases there are also elements that seem to suggest a belief in persons as separately existing entities.[7] And it is not difficult to see why clear-cut instances of this position might be hard to come by, since if it is true that persons exist and it is also the case that the concepts of person and personal identity are simple and unanalyzable, we should expect there to be some distinct entity the existence and persistence of which explain these singular facts about persons and their identity over time. The Vātsīputrīyas or Pudgalavādins of the Buddhist tradition do, though, represent a clear instance of this type of Non-Reductionism; this is precisely the force of their claim that the person (pudgala) is, while existent, neither identical with nor distinct from the skandhas.[8]

Parfit does not himself discuss Eliminativism, so to see where this fits on the philosophical terrain we need to turn to those authors, namely Giles and Stone, who advance this view as the Buddhist response to Parfit's Reductionism. Unfortunately, Giles is not especially helpful, apparently falling prey to some serious terminological confusion. He begins by distinguishing between reductionism and eliminativism through the use of an example from philosophy of mind, the dispute between identity theorists and eliminative materialists. And he notes the characteristically eliminativist tenor of the latter's claim that folk psychology is so hopelessly confused as to defy translation into any scientific discourse, and so should be abandoned. Yet he goes on to propose that we call Eliminativist the early Buddhist anātman theory (as well as Hume's *Treatise* position), despite the fact that on that view we will, "on pragmatic grounds, continue to permit the use of the language of personal identity" (1993, p. 176). But if talk of persons and their identity over time has some utility (as all Buddhists in fact maintain), then our theory of persons cannot be a candidate for utter elimination. He also takes the *Milindapañha* denial that the person is identical with the psychophysical complex as a whole to be a rejection of Reductionism. But Reductionists do not claim that a person is identical with a certain sum of impersonal elements; instead they claim that the existence of a person just is the occurrence of certain impersonal elements. While Giles sees the importance of the Buddhist theory of two truths, he fails to see how it may be used to mark the difference between "is identical with" and "just is."

Stone (1988) is more helpful. Eliminativism is to be distinguished from Reductionism not in terms of the denial of a self (both agree that we are not "something extra") or in terms of the denial that persons are to be found in our ultimate ontology (both deny that persons have this privileged status), but rather in terms of the question whether the attitudes we ordinarily take toward ourselves and others are at all coherent. The Non-Reductionist claims that such things as prudential concern, anticipation, regret, responsibility for past deeds, merit, and the like all require that there be something extra, over and above body and brain, and so forth. The Reductionist denies that there is this something extra, but holds that such attitudes may still be rational (even if their scope is somewhat altered when we come to accept Reductionism). The Eliminativist agrees with the Reductionist that we are not something extra, but also agrees with the Non-Reductionist that our attitudes toward persons are coherent only if we are something extra. Prudential concern, hopes, fears and regrets, judgments of responsibility, merit, and praise and blame—all these are irrational. And since Locke is right to see the forensic elements as central to the concept of a person, it follows that all talk of persons is deeply incoherent. In place of the mildly dismissive Reductionist attitude toward persons—as "mere constructions" out of more fundamental entities—we find in Eliminativism an outright rejection of all that persons are thought to be. Like the demons believed in by our ancestors, persons are posits of an utterly misguided theory.

Like Giles, Stone identifies the Buddha as an Eliminativist. As will already be evident, I believe this is mistaken. But it may prove worthwhile to examine why some might see early Buddhism and Abhidharma as Eliminativist rather than Reductionist. First we need to see how the taxonomy we have developed so far might need adjustment in order to accommodate the Buddhist problematic, and so we turn to an examination of some representative early Buddhist and Abhidharma texts. When this is done we will return to the question of reading Buddhism as Eliminativist.

One passage in the early Buddhist text *Milindapañha*[9] clearly exhibits signs of Reductionism. The text as a whole is in the form of a dialogue between the Buddhist monk Nāgasena and a king, Milinda by name, who seeks to enhance his understanding of Buddhism by asking Nāgasena a series of probing questions concerning the system. The passage in question opens with Milinda asking whether adult and infant are the same person or distinct persons. Nāgasena replies that they are neither the same person nor distinct persons,[10] and asks Milinda's view. The king replies that adult and infant are distinct persons. This answer is rather surprising, since throughout the work Milinda tends to represent the commonsense view of things, and most people would judge the adult Milinda to be the same person as the infant Milinda. But it appears from

the context as if in this case the king is not to serve as representative of our commonsense intuitions, but is rather expressing a certain understanding of the Buddha's teachings on persons. The king has already learned through his discussions with Nāgasena that the person is made up of five skandhas, none of which exists continuously throughout the course of a lifetime, and thus that the person is devoid of a self. He now reasons that since none of the skandhas that made up the infant is present in the adult, adult and infant must be distinct persons. He supposes, that is, that personal identity over time requires the continued existence of some one entity through the distinct stages in the life of a person; since the skandha analysis reveals the absence of any such entity, it follows that personal identity does not extend over any substantial portion of a lifetime.

Anyone who has taught the early Buddhist argument for nonself (anātman) from the impermanence of the skandhas will recognize this response, since it is quite common for students to understand the Buddhist teaching in just this way. But Nāgasena makes it clear through a series of reductios that Milinda has reasoned incorrectly. By the same reasoning, it would follow that there are no mothers or fathers, no educated persons, and no one who deserves punishment for past crimes. A mother, for instance, is a person who conceives, bears to term, and then gives birth. But the skandhas making up the woman who conceives are no longer present, for example, in the woman carrying a second-trimester fetus. By Milinda's reasoning, then, the woman who gives birth is not the same person as the woman who conceived, or the woman who bore the fetus, and so is not a mother. Milinda agrees that these results are unacceptable, and asks Nāgasena how he views the matter. Nāgasena replies that adult and infant are the same person. This would appear to contradict his earlier response, that adult and infant are neither the same person nor distinct persons. But Nāgasena goes on to explain that those skandhas making up the adult have as their causal antecedents the skandhas that made up the infant; impermanent elements existing at distinct times are collected together—that is, make up a person—when they bear the right sorts of causal relations to one another.

This is illustrated with the example of the one light that shone all night. If a lamp were to be lit in the evening and burned continuously until the morning, we would agree that there was one light—one source of illumination—that shone all night. Yet it is agreed that that which actually illuminates at any one moment during the night, namely a flame, is numerically distinct from that which illuminates at any other moment. This follows from the physics of flames, for a flame is a collection of fire atoms (by classical Indian physics) or incandescent hydrocarbon molecules (by our physics), these entities undergoing constant replacement. Yet we do not, for all that, say that there were many distinct

lights illuminating the room over the course of the night; it is not incorrect to say that there was just one light shining all night. This, explains Nāgasena, is because the flame at 4:00 A.M. has as its remote causal antecedent the very first flame in the series, that which occurred when the lamp was lit at 9:00 P.M. As long as the right conditions obtain, a given collection of fire atoms will, in going out of existence, cause a new collection to come into existence. Whenever distinct flames form such a causal series, this may be referred to as one continuously existing light.

Earlier, Nāgasena told Milinda that the name "Nāgasena" is a mere convenient designation for a causal series of sets of impermanent skandhas. So here he might say that the expression "one light that shone all night" is a convenient designation for a causal series of collections of fire atoms. In the *Nikāyas*, the canonical literature of early Buddhism, "chariot" was the stock example of a convenient designation. In the Abhidharma literature we find such other instances as "forest," "village," "oven," "head hair," and "army." To call "forest" a convenient designation is to say approximately the following: given the regular occurrence in the world of certain sorts of clusterings of trees, and the nature of the interests we have in such clusterings, it has proven advantageous that we have a single term that may be used to refer to such clusterings, and thus avoid the prolixity involved in referring to the individual trees and their relations to one another.

The concept of a convenient designation plays a key role in a two-part strategy for undermining belief in a self. First it is argued, through the use of the skandha analysis, that there is no empirical evidence for the existence of an entity having the properties of a self: continued existence throughout a lifetime, being the subject of experience, performing the executive function, and the like. Then it is argued that our belief in a self is generated by our use of such convenient designations as "person," the personal pronouns, and personal names, in conjunction with the misguided acceptance of a naive semantic realism that takes the meaningfulness of a word to require the existence of some entity bearing that word as its name. Thus, since all of our ways of conceptualizing persons may be accounted for without supposing there to be anything more to the existence of a person than just a complex causal series of impermanent collections of skandhas, and there is no empirical evidence for the existence of anything other than the skandhas, we have no reason to believe that the existence of a person involves anything other than impersonal phenomena in a complex causal series.

It is possible to see the strategy that Parfit employs in defense of Reductionism as proceeding along similar lines. For Parfit first argues (1984, pp. 223–228) that we do not have the sort of evidence that would support the claim that we are separately existing entities. He then proceeds to try to show, through the use of various puzzle cases, that per-

sonal identity over time is subject to the same sorts of sorites difficulties as are heaps, clubs, and nations. And since most of us are willing, when confronted with such difficulties, to say that heaps, clubs, and nations are said to exist only because of the way that we talk, this suggests that Parfit sees "person" as a kind of convenient designation. If such a reading is correct, then this parallel counts as evidence in support of the claim that early Buddhism and Abhidharma are Reductionist. But Nāgasena's use of the notion of a convenient designation brings out something else as well. While the *Milindapañha* does not employ the doctrine of the two truths, this passage nicely illustrates the tensions that led to the development and articulation of this device in the Abhidharma.

The doctrine of the two truths distinguishes between what is called conventional (saṃvṛti) truth and ultimate (paramārtha) truth. A statement is said to be conventionally true if it conforms to common sense, that is, if it is in accordance with conventionally accepted linguistic and epistemic practices. If we use "conceptual fiction" to refer to whatever is thought to exist only because of our use of a convenient designation, then "ultimate truth" may be defined as follows: a statement is ultimately true if and only if it corresponds to the facts and neither asserts nor presupposes that conceptual fictions exist. Given a sufficiently restrictive ultimate ontology, it will then turn out that most statements that are conventionally true are ultimately false. It thus becomes necessary to explain why most such conventionally true but ultimately false statements appear to have utility for human practice. The explanation is simply that while certain of the entities quantified over in such a statement do not exist, it is possible systematically to replace all talk of such entities with talk of entities that do ultimately exist, thereby arriving at a statement that is ultimately true, that is, that does correspond to facts the constituents of which belong to our ultimate ontology. Thus, most conventionally true but ultimately false statements are amenable to full translation[11] into ultimately true statements; all that is lost in such translation is the misleading implication that conceptual fictions exist.[12]

This device allows us to resolve the seeming contradictions in Nāgasena's position. When he claims that adult and infant are neither the same person nor distinct persons, he is stating what he takes to be an ultimate truth. Since persons are conceptual fictions, any claim concerning the identity over time of a person must be ultimately false. His assertion that adult and infant are the same person he takes to be only conventionally true. Likewise Milinda's assertion that adult and infant are distinct persons Nāgasena claims to be conventionally false. For it is our conventional practice to refer to infancy and adulthood as merely two stages in the life of one continuous person. And a whole host of other customary practices involves this as presupposition, for instance the notion that as adults we are obligated to care for our aged parents

because they cared for us as infants. But the conventional truth that I am the same person as the child pictured in a certain old photograph may be accounted for in terms of facts about a complex causal series of collections of skandhas, just as the conventional truth that there was one light that shone all night may be accounted for in terms of the facts about a causal series of collections of fire atoms. We may eliminate all talk of such conceptual fictions as persons and enduring lights, yet preserve the underlying truths that such statements are attempting, in their rough-and-ready way, to assert.

I think we can now see why early Buddhism and Abhidharma have often mistakenly been seen as Eliminativist. The error arises through attending solely to what is said at the ultimate level of truth, and failing to appreciate the relation between the ultimate and conventional levels of truth. Eliminativism is not simply the view that talk of persons may in principle be eliminated. Both Reductionist and Eliminativist maintain that ultimately there are no persons. But the Eliminativist urges in addition that the claim that there are persons be seen as conventionally false as well, since the Eliminativist maintains that our commonsense theory of persons is incoherent, or at least so misleading as to be more troubling and confusing than theoretically useful.[13] By contrast, the Reductionist holds that while unquestioning adherence to the commonsense theory of persons does result in misguided views about how we should live our lives, the theory does have its uses, which fact requires explanation; hence it is conventionally true, though ultimately false, that there are persons. Like Milinda's wrong view about adult and infant, the reading of early Buddhism as Eliminativist results from the failure to consider the semantic dimension of the dispute.

I claimed earlier that a reductionist about things of kind K sees much of our talk of Ks as *not wholly incorrect.* The doctrine of the two truths gives us one way of understanding what this "not wholly incorrect" might come to: ultimately false, since "K" is a mere convenient designation and Ks are not in our ultimate ontology; but still conventionally true, since talk of Ks both has utility given our interests and customary practices, and may systematically be replaced with talk of entities that are in our ultimate ontology. There are those, however, who, while describing themselves as reductionists, would resist this assimilation, since they have serious reservations about the notion of an ultimate ontology. This is clearly so for those so-called "weak" Reductionists, such as Shoemaker and Brennan, who deny that talk of persons is eliminable from our discourse, and deny that persons are logically constructed out of more particular and completely impersonal things. But this might also be true of a "strong" Reductionist like Parfit, who seems somewhat reluctant to commit himself to any one ultimate ontology.[14]

Now such reluctance may reflect nothing more than a rhetorical

strategy designed by Parfit to appeal to the widest possible audience. But it might instead come from the belief that any ontology must always be provisional, forever open to revision in the light of future experience and future changes in human interests and practices. I would maintain, however, that reductionism requires minimally that the notion of an ultimate ontology be held to be coherent, something to whose attainment we may at least sensibly aspire. For the normative force of reductionist claims requires ontological backing.[15] The reductionist's strategy is to persuade us that Ks have less importance than we are wont to believe, by first convincing us that Ks are really just constructions on more primitive entities. And the notion of relative primitiveness at work here makes sense only on the assumption that there is such a thing as the ultimate ontology. Parfit first argues that we are not what we believe (separately existing entities), and then uses this result to support his claims concerning rationality and morality.

The effectiveness of such a strategy depends on the implicit premise that persons would not have the sort of rational and moral significance ordinarily ascribed to them unless they were themselves ultimately real entities. Indeed, the theory of two truths, and the view of ontology that that theory implies, are admirably suited to express the points that Parfit seeks to express—for instance, by claiming that the facts of personal identity over time and the separateness of persons are "less deep" on the Reductionist view, and so lack the rational and moral significance we tend to give them (e.g., 1984, p. 337). For the metaphor of depth may be replaced along the following lines: a fact is "less deep" just in case a statement expressing that fact is conventionally but not ultimately true. But to say this is just to say that some of the entities referred to in that statement are not ultimately real, are not to be found in our ultimate ontology.

That the account of persons developed in early Buddhism and Abhidharma is meant to have normative force is clear. Liberation from the cycle of rebirth is said to result from the realization that there is no self, and that the continued existence of a person just consists in a complex causal series of collections of physical and psychological elements. And liberation involves the rejection of that mode of life typical of the "householder," with its characteristic attachments to home, family, and occupation as ongoing enterprises, as well as its characteristic concern for one's prospects upon rebirth. All such attachments and concerns are to be replaced by a way of life marked by equanimity, spontaneity, and the developed capacity to feel sympathetic joy at the welfare of all. The sorts of significance that we ordinarily attach to our life projects and our situation beyond this life are undermined by the truth about what we ultimately are.

Buddhism is often said to be a kind of middle path between two

extremes. A number of different pairs of opposing extreme views are identified in the tradition, but one such pair is frequently singled out for special emphasis, namely that of eternalism and annihilationism. Eternalism is the view that there is an eternal self, and thus that rebirth is transmigration. Annihilationism is usually portrayed as the view that the self ceases to exist at the end of a single lifetime. Eternalism is said to have the normative consequence that since one will deserve the karmic fruits reaped in the next life from one's present deeds, we all have a reason to act in accordance with the karmic moral rules. Annihilationism, by contrast, is said to result in a radical antinomianism: since there can be no karmic retribution beyond this present life, one has no reason to act morally where doing so involves sacrificing one's own immediate gratification. Both views are, the Buddhist claims, false because of their shared presupposition that a self exists. Yet some of the normative consequences of eternalism must be preserved, since belief in the karmic moral order is required if persons are to progress toward enlightenment.

The middle path between eternalism and annihilationism thus consists of a demonstration that rebirth is compatible with the nonexistence of the self. Since the continued existence of a person in one life just consists in the obtaining of appropriate causal connections among various physical and psychological events, the continued existence of a person over several lives is likewise possible in the absence of an enduring self, provided the right sorts of causal connections obtain between lives. And our alleged ability to recall events from past lives presumably shows that such connections do obtain. Thus to the extent that one is justified in feeling concern for what happens to oneself in the later stages of this life, concern is equally justified with respect to one's future lives. Annihilationism wrongly assumes that only the continued existence of a self could give one a reason for self-interested concern, yet this is clearly false in the case of a single lifetime. As the life of the enlightened person demonstrates, one can know that there is no self yet not lose oneself in a "solipsism of the present moment."[16] While enlightened persons do not exhibit self-interested concern in the same way and to the same extent as the unenlightened, they do appear to be motivated by considerations concerning how their present acts will affect them in the future.

Parfit describes three possible views about what it is rational for an agent to seek: (1) the classical self-interest theory (S), according to which we as rational agents should ultimately be governed by a temporally neutral bias in our own favor; (2) the extreme claim (E), that if Reductionism is true then we have no reason to be concerned about our own futures; and (3) the moderate claim (M), that if Reductionism is true, then the causal connections obtaining between different stages in our lives give us some reason to be concerned about our own futures. Parfit

claims that since Reductionism is true, S is false; but he also asserts that both E and M are defensible.[17] Clearly, this dispute does not map perfectly onto the Buddhist problematic of eternalism, annihilationism, and a middle path between them. I would claim, however (though I cannot argue at any length for this here), that the Buddhist discussion of eternalism and annihilationism suggests that M should be accepted and E rejected. Like annihilationism, E appears to be motivated not by Reductionism but by Eliminativism. The Reductionist would agree that since there are ultimately no persons, such anticipatory attitudes as dread and such retrospective attitudes as regret cannot rationally be justified at the ultimate level of truth.[18] But persons are conventionally real: the practice of speaking of ourselves as persons has greater overall utility than the available alternatives.[19] And so, certain person-involving attitudes may turn out to be rationally justifiable at the conventional level of truth.

Consider this analogy. Most of us would agree that a city just consists in certain buildings, streets, persons, and so forth, arranged in certain characteristic ways; strictly speaking there are no cities. Still, given the utility of the convenient designation "city," a certain degree of civic pride may be justifiable. The overweening pride of civic chauvinism is ruled out, since it seems to require that one think of the city as a separately existing entity. But urban aggregations themselves have some degree of utility, and this is enhanced by the behavior that results when their inhabitants exhibit some appreciation for the character of the particular aggregate in which they reside. Thus the practice of encouraging civic pride has a consequentialist justification. By the same token, the practice of thinking of ourselves as persons can be expected to have significant utility. Much of this stems from its facilitating such person-involving attitudes as anticipation and regret. Suppose that I, realizing that the action I now contemplate will result in future pain for me, am deterred through anticipating that I shall experience pain. Now there is no further fact that makes it the case that the person who will feel that pain will be me; this fact just consists in the obtaining of certain relations among certain purely impersonal present and future entities and events. I do nonetheless have a special reason for refraining from the action, namely the fact that this is (typically) the best way to insure that that future pain does not occur. Because pain is bad, we all have a reason to try to prevent its occurrence; and in general I am better positioned than anyone else to prevent my own future pains. Existential dread may be unjustifiable, but a moderate degree of concern over one's anticipated future pain does have considerable utility.

Buddhist Reductionism thus has significant normative consequences. I claimed above that such consequences require ontological grounding, hence that reductionism is to be construed as a thesis about our ultimate ontology. It would be fruitful to examine how the Buddhist Reductionist

proposes to determine the contents of our ultimate ontology. It is striking, for instance, that while sorites-induced difficulties are crucial to Parfit's argument against further-fact Non-Reductionism, sorites paradoxes play no role whatever in the arguments used to support Buddhist Reductionism. Indeed, the Indian tradition overall has no Chariot of Devadatta problem to answer the Ship of Theseus puzzle. Instead, in the texts of early Buddhism there is the expectation that we will simply agree that chariots are ultimately unreal, once it is pointed out that a chariot is an assemblage of parts. In later Abhidharma texts we can discern an argument against the existence of wholes in general, but this argument makes no use of the boundary-setting difficulties that the admission of partite entities presents us with. I would suggest, though, that this is not because Indian philosophers were simply ignorant of sorites phenomena. Rather, those Indian philosophers, the Naiyāyikas, who championed the existence of wholes went to extraordinary lengths precisely to insulate their doctrine of complex substances from problems of indeterminacy.

Thus, Nyāya maintains that the addition or subtraction of a part from a given substance results in the destruction of that substance and the coming into existence of a new substance. And likewise for change in the arrangement of the parts, and for such qualitative changes as change of color, taste, odor, texture, and the like.[20] This position obviously rules out many of the sorts of spectra of indeterminacy that would otherwise plague the champion of complex substances. I know of no Nyāya text explicitly linking this position with the intention to avoid such difficulties. But I find it hard to imagine what else might motivate the adoption of such a counterintuitive view, particularly on the part of the eminently sensible Naiyāyikas. In any event, the Buddhist Reductionist was not given the opportunity to exploit sorites difficulties in arguing against persons and other such wholes; those difficulties had been anticipated by the opponent.

Now it would still be open to the Buddhist Reductionist to use sorites difficulties to argue against the existence of persons, since if these are to explain personal identity over time then they must be thought of as complex substances that endure through replacement of parts, qualitative change, and the like. (Nyāya accounts for personal identity over time by means of the continued existence of an impartite self, not through the continued existence of a person.) But rather than pursuing this strategy, they seek instead to show that partite entities in general must be thought of as mental constructions and are thus not ultimately real. But then the ultimate point of the sorites strategy as wielded by the reductionist about Ks is precisely to show that the Ks in question are just the result of human conventions, are a mere mental fabrication.[21] For genuine vagueness induces bivalence failure at many different levels: not only is there a gray area where one may say of a given number of grains

of sand neither that it is nor that it is not a heap; it is also indeterminate just where this zone of indeterminacy begins and ends in the heap spectrum.

Genuine vagueness is thus not the result of mere ignorance, something that could always be overcome by one sort or another of precisification. And for the metaphysical realist, it is inadmissible that mind-independent reality should in itself be the source of such indeterminacies. If mind-independent reality is to be the final arbiter of the truth of propositions, then reality itself must be fully determinate; any failures of bivalence in our theory about the world can only result from that theory's not having carved up reality at its joints—from our having used concepts and categories not wholly derived from the nature of mind-independent reality. To the metaphysical realist, sorites phenomena can only indicate an element of mental construction; it is this that makes such phenomena useful for reductionist purposes. The Buddhist Reductionist simply cuts to the chase—avoids discussing sorites phenomena and proceeds directly to the notion of mental construction.

The Buddhist Reductionist argument against the existence of real wholes—that is, for the conclusion that the partite is mentally constructed—is relatively simple and straightforward. Suppose we agree that the parts of, for example, the chariot are themselves real. If, in addition to the chariot parts, the chariot itself is thought to be real, then it cannot be said to exist distinct from the parts given that these are related to one another in the manner that results from their assembly. For there is no evidence for the existence of a chariot that is not just evidence for the existence of one or more chariot parts and their assembly relations; a distinct chariot is a superfluous posit. But neither can it be said that a real chariot is identical with the assembled chariot parts. For the chariot may be said to have n parts, while the assembled parts may not be said to *have* n parts—they can only be said to *be n in number.* And if both chariot and parts are real, then the chariot must be either identical with or distinct from its parts. Those such as the Pudgalavādins who claim that the whole is neither identical with nor distinct from the parts are easily convicted of logical incoherence.[22] There still remains the possibility that the chariot is real while its parts are unreal. But the chariot is itself a part of a larger whole, namely the universe; if only wholes are real, there can only be one real thing. There would then arise seemingly insuperable difficulties in trying to account for the apparent utility of the myriad distinctions we routinely draw. We must conclude that only impartite entities are ultimately real; the partite is mentally constructed out of those impartite entities that regularly co-occur in ways that have a high degree of saliency for sentient systems like ourselves.

This is the reasoning behind the development of the various dharma theories of the Abhidharma schools. The dharmas are those entities that

are contained in our ultimate ontology; and the argument tells us that they must be impartite. But this does not result in the sort of simple-minded Democritean atomism that one might expect. The universally agreed upon definition of a dharma is: that which bears its own essential nature (svabhāva). The basic thinking here is that all partite entities must borrow their essential properties from parts: the characteristic shape of the chariot is a function of the shapes of its parts, its utility as a means of transport is a function of various interrelations among its parts, and so forth. Thus, whatever does not borrow its essential properties from other things, but instead bears that nature as its own, must be ultimately real and not a mere mental construction. This approach has some important advantages. It gives a useful way to extend the partite/impartite distinction beyond what is spatially extended to those psychological events that make up the so-called nāma skandhas. It thus becomes possible to think of a particular pain sensation, for instance, as impartite, and a complex emotional episode like an occurrence of jealousy as partite. It likewise opens the way to analyzing physical objects not into aggregates of indivisible atoms with determinate size, but rather into bundles of simple property particulars.[23]

Most importantly though, the svabhāva criterion of dharma-hood yields a uniform way of telling whether something is ultimately real or merely a mental construction: is it analytically findable, or does it dissolve upon analysis? Thus Buddhaghosha tells us that head hairs are mere conceptual fictions since they can be analyzed into color, shape, solidity, and smell; a particular smell, on the other hand, presumably remains as the terminus of any analysis, and is thus a dharma.[24] To borrow a Russellian distinction, conceptual fictions can be known by description, whereas dharmas can only be known by acquaintance. A description, of course, utilizes a combination of two or more concepts. Thus an element of aggregation must enter into the constitution of that which is represented as one thing yet can be known by description. And combination or aggregation is a mental contribution; all that is ever actually given in experience is co-occurrence. In pure acquaintance, on the other hand, the mind is wholly passive and receptive to the given. That which bears its own essential nature is free of all taint of mental construction; that which borrows its essential nature from what is other can only be a conceptual fiction, the product of an inveterate tendency of the mind to construct unreal aggregates.

This view depends, in the end, on the soundness of the Abhidharma argument described above for the conclusion that wholes are unreal. And some find that argument unpersuasive. It should be recalled, though, that sorites arguments can always be constructed to try to show that a given sort of partite entity is conceptually constructed. And there is some reason to believe that such attempts will generally be successful. It is, for

instance, relatively easy to show that complex organisms cannot be ultimately real, given the many indeterminacies that arise in connection with their continued existence as their parts are replaced. There is also a problem in supposing that a given organism has a determinate species membership, given that our species concepts are themselves fluid and responsive to a variety of pragmatic pressures. Indeed, philosophical discussions of "natural kind" terms tend to overlook the actual practice of biologists, chemists, and physicists in their use of these classificatory devices—not to describe unalterable regularities in nature, but to express those idealizations of observed tendencies that are necessary for purposes of theory construction. The actual data of the natural sciences provide fuel for countless sorites arguments against the ultimate reality of the various entities posited by mature sciences. Possible-worlds machinery yields another rich vein of sorites arguments, since this allows us to construct spectra of indeterminacy across worlds for any partite entity.

Thus there appears to be some reason for the Reductionist to embrace the sort of ultimate ontology described in the Abhidharma dharma theories. There is, though, one final step in the development of Buddhist Reductionist thinking about our ultimate ontology. This step seems not to have been taken by any of the classical Abhidharma schools, but only appears with the rise of Yogācāra-Sautrāntika in the work of Diṅnāga. It is the claim that what is ultimately real is just the svalakṣana, the ineffable pure particular. This radical nominalism results from the realization that the same considerations that militate against the existence of partite entities apply with equal force to allegedly real universals and resemblances. The relation between a universal and its instances turns out to be just as problematic as that between a whole and its parts. And the universal is just as much the product of the mind's tendency to posit a one when it has collected together the many. Of course the Buddhist nominalist is faced with the daunting task of explaining the efficacy of our discursive practices, given that at least some of the terms of a language must be general.[25] But the Yogācāra-Sautrāntika responds to this challenge with the theory of apoha, according to which the meaning of "cow" is "not non-cow." Here the "not" is to be read as exclusion negation and the "non" as choice negation. The resulting formal model, and the psychological machinery that is said to instantiate the model, are together designed to explain how a speaker can learn to use "cow" to refer to just those particulars that other speakers of the language agree in calling cows, given that there is no cowness universal inhering in, or real resemblance shared by, those particulars.

At this point, those with Reductionist inclinations might begin to balk at the price being asked of them: an ultimate ontology of ineffably unique pure particulars, a two-level truth theory, and a radically nomi-

nalistic semantics. It must be borne in mind, though, that Reductionism is espoused because of its normative consequences, which require ontological grounding. And the notion of ontological grounding at work here is in turn based on metaphysical realist presuppositions. Coming to know the truth about what we are is supposed to change our views about how to live our lives precisely because we take "truth" to be correspondence to mind-independent reality. We will be inclined to transform our habitual modes of conduct only to the extent that we find them to stem from beliefs contaminated by elements of mental construction—by styles of thought that leave the mind free to project its desires onto the world. Reductionists must be metaphysical realists, and as metaphysical realists must seek to purge all elements of mental construction from their ultimate ontology. Buddhist Reductionism is consistent and complete Reductionism.

NOTES

An earlier version of this essay was read at the East-West Philosophers' Conference of the Australasian Society for Asian and Comparative Philosophy, held at Massey University, Palmerston North, New Zealand, 12–15 August 1994. I wish to thank the following for helpful comments on prior drafts: David L. Anderson, Kenton Machina, Roy Perrett, C. Ram-Prasad, and Liane Stillwell.

1 – See Stone 1988 and Giles 1993. Parfit now calls the Buddha's view "Eliminative Reductionism," which he distinguishes from his own view, now called "Constitutive Reductionism" (1995, pp. 16 f). For reasons that will become evident below, although I once described the view of early Buddhism and Abhidharma as a form of eliminative reductionism, I now see this as a mistake.

2 – Duerlinger (1993) represents both Vātsīputrīya and Prāsaṅgika Madhyamaka as holding Non-Reductionist positions. His evident sympathy for Candrakīrti's views would seem to suggest that he takes at least the Prāsaṅgika position on persons to be a defensible interpretation of the Buddha's view.

3 – I shall use "Reductionism" to refer to the view of persons and personal identity developed in Parfit 1984, and "reductionism" to refer to any theory that attempts to reduce entities of one sort to entities of a distinct sort. As shall soon become clear, not every reductionist about persons is a Reductionist.

4 – It might be thought that this commitment to the possibility of a completely "im-mob-ish" description places a severe constraint on

the reductionist, in that such a description will necessarily omit just those elements that distinguish mob behavior from the behavior of individuals and small groups. But this is not so. Once we know all the facts about all the thoughts, feelings, and actions of each individual, we will have captured all the ways in which the actions of each affected the behavior of other individuals in the mob.

5 – At least not on any known version of the demonic-possession theory. If, however, such a theory had posited distinct types of demons for viral and bacterial pneumonia independently of the development of the germ theory, we might be more inclined to look for translation rules.

6 – Noncircularity is obtained through the use of a "Ramsey sentence" approach. See Shoemaker and Swinburne 1984, pp. 99 f.

7 – Swinburne (Shoemaker and Swinburne 1984, p. 27) represents all three authors as further-fact theorists. Noonan (1989, p. 19) takes them as representative of what he calls the Simple View, which is a form of the separately-existing-entity thesis.

8 – See *Kathāvatthu,* pp. 1–71, *Abhidharmakośabhāṣya* 9, pp. 462–463.

9 – While the *Milindapañha* appears to have been written during the period of Buddhist scholasticism (the Abhidharma period), and thus not during that phase of the Buddhist tradition commonly identified as the era of early Buddhism, the text appears sufficiently free of commitment to partisan Abhidharma scholastic positions as to seem better classified as early Buddhist. The episode in question is at II.2.1.

10 – This is the most natural reading of the Pāli, namely as denying both numerical identity and numerical distinctness, qua persons, of adult and infant. This and similar passages are, however, sometimes translated more simply by the phrase "neither the same nor different," which is then interpreted (e.g., by Collins [1982]) as involving an equivocation: neither qualitatively identical nor numerically distinct. Such a reading misses the connection between these passages and the Buddha's treatment of the "indeterminate questions" (avyākṛta), which he compares to such questions as, when a fire has gone out, in which of the four possible directions it has gone. One may sensibly deny that the fire has gone in any of these directions without equivocating on "gone to the north," and so forth. On the avyākṛta as involving bivalence failure, see Ruegg 1977.

11 – The notion of full translatability is somewhat problematic, since the full expression, in the privileged discourse, of certain conventional

truths would seem to require strings of indefinite if not infinite length, and it is far from clear that such a string could count as a translation. In defense of the claim of full translatability, it could be said that our inability to produce or comprehend indefinitely long strings is a "mere practical difficulty," that is, only reflects human limitations and so does not represent a shortcoming in the analysis itself.

12 – It might appear unduly harsh to call all statements employing convenient designations false, rather than just potentially misleading. The Buddhist reductionist agrees that such convenient designations as "house" and "forest" are relatively innocuous. Our use of the term "person," though, regularly leads to the most dire consequences. Parfit makes a similar point about the term: "Though we need concepts to discuss reality, we sometimes confuse the two. We mistake conceptual facts for facts about reality. And in the case of certain concepts, those that are most loaded with emotional or moral significance, we can be led seriously astray. Of these loaded concepts, that of our own identity is, perhaps, the most misleading" (1995, p. 45). Parfit merely refrains from taking the next step: declaring all employment of the concept useful but false.

13 – Of course the Eliminativist cannot claim that it is conventionally false that there are persons, since the theory of persons is, as things now stand, accepted by common sense. Eliminativists can only advocate that the "folk theory" of persons be excised from our common sense, and that our language be correspondingly revised. They do so on the grounds that acceptance of the theory leads to beliefs, desires, and actions that presuppose we are something extra; since we are not something extra, practical reason informed by this theory will inevitably encounter insuperable difficulties. See Stone 1988, pp. 525–530.

14 – This reluctance might be another factor that has led some to see sufficiently great differences between Parfit's Reductionism and the early Buddhist position as to classify the latter as Eliminativist. Early Buddhism (as well as Abhidharma) seems quite prepared to say what should go in our ultimate ontology—and that persons are not to be found there. This is not tantamount to Eliminativism, but the apparent contrast with Parfit's view might have led some to see it as such.

15 – It might be wondered whether, for example, the reduction of organic chemistry to quantum mechanics actually carries any normative weight. But the relative status of the disciplines of physics and chemistry seem to derive at least in part from the possibility of

such reduction. The prestige of physics appears to stem in some measure from this notion that it approaches more closely to a grasp of reality undistorted by the limitations imposed by such factors as the scale of our sensory apparatus or our ability to track only so many particulars at once.

16 – Here is further evidence that the Extreme view E (see below) is inconsistent with Reductionism: the Extremist seems to hold that while I have no reason for concern about my future, I do have reason for concern about my present state. Presumably this is because that future person will not be me, but the present person is me. But it is ultimately false that that future person is not me and that the present person is me. Ultimately there are no persons. If there are any facts that explain the rationality of my concern for persons, these facts are only visible at the conventional level of truth.

17 – Parfit's view may have changed. He writes (1986, pp. 836–837) that while it is, on the Reductionist view, irrational to view personal identity as what matters, it is rational to care about the causal connections obtaining between, say, a person and that person's Replica. This would appear to rule out E as well as S. He does not, however, give any argument against E.

18 – Stone's (1988) argument for Eliminativism depends crucially on the claim that such attitudes are incoherent in the absence of the further fact that Non-Reductionism requires, hence that our theory of persons is incoherent given that there is no such further fact.

19 – For the Buddhist Reductionist the utility of the conceptual fiction of persons is, in the final analysis, soteriological: it is because we think of ourselves as persons that we seek to avoid suffering and thus are disposed to enter the Path to nirvāṇa. See Collins 1982, p. 152. In this respect Buddhist Reductionism resembles those Indian Non-Reductionist systems (such as Sāṃkhya and Advaita Vedānta) that take liberation from suffering as the primary goal. Such systems commonly posit knowledge of self as necessary for liberation, but then worry that the self, as subject of knowledge, cannot take itself as object of knowledge. The paradox is resolved by claiming that the self comes to know itself indirectly, namely by overcoming the error of misidentifying with some inappropriate category. Ignorance thus plays a crucial role in the soteriologies of these systems as well.

20 – See Ramaiah 1978, pp. 61–90.

21 – So Horgan (1994), for instance, after arguing that genuine vagueness is logically incoherent and hence impossible, seeks to reconcile this with the evidence for widespread vagueness in the empirical world,

by claiming that while vagueness is impossible in mind-independent reality, it actually occurs in thought and language, where it possesses some utility and where the deleterious effects of its logical incoherence may be contained through various insulating strategies. This position seems to stem from a view that has long been advanced by Kenton Machina.

22 – See *Katthāvatthu,* pp. 3–71. Also see Ruegg 1977, pp. 34–36.

23 – This would appear to have been the tack taken by certain Sautrāntikas in their account of the rūpa dharmas. See *Vijñaptimātratāsiddhi* 14, where the opponent is a realist who rejects atomism yet holds that there are rūpa dharmas. Vasubandhu's arguments against this position in verse 15 make clear that the view in question analyzes physical objects into bundles of property-particulars devoid of substrate.

24 – *Visuddhimagga* XI.88.

25 – Indeed given that the svalakṣana or pure particular is radically momentary, all the terms of a language will have to be general. For then a svalakṣana is simply too evanescent to be dubbed. What we actually succeed in referring to through, for example, the use of a demonstrative, is always a series of "resembling" particulars. There can be no "logically proper names."

REFERENCES

Abhidharmakośabhāṣyam of Vasubandhu. 1975. Edited by Prahlad Pradhan. Patna: K. P. Jayaswal Research Institute.

Brennan, Andrew. 1988. *Conditions of Identity: A Study of Identity and Survival.* New York: Oxford University Press.

Collins, Steven. 1982. *Selfless Persons.* Cambridge: Cambridge University Press.

Duerlinger, James. 1993. "Reductionist and Nonreductionist Theories of Persons in Indian Buddhist Philosophy." *Journal of Indian Philosophy* 21 : 79–101.

Giles, James. 1993. "The No-Self Theory: Hume, Buddhism, and Personal Identity." *Philosophy East and West* 43 : 175–200.

Horgan, Terrence. 1994. "Robust Vagueness and the Forced-March Sorites Paradox." *Philosophical Perspectives* 8, edited by James E. Tomberlin (Atascadero, California: Ridgeview): 159–188.

Kathāvatthu. 1961. Edited by Kassapa, Bhikkhu Jagadosa. Bihar Government: Pali Publication Board. Translated as *Points of Controversy* by T. W. Rhys Davids. London: Pali Text Society, 1915.

Milindapañho. 1972. Edited by R. D. Vadekar. Bombay: Bombay University Publications. Translated as *The Questions of King Milinda* by T. W. Rhys Davids. Sacred Books of the East, vols. 35, 36. Oxford: Oxford University Press, 1890.

Noonan, Harold. 1989. *Personal Identity.* London: Routledge.

Parfit, Derek. 1984. *Reasons and Persons.* New York: Oxford University Press.

———. 1986. "Comments." *Ethics* 96 : 832–872.

———. 1995. "The Unimportance of Identity." In *Identity,* edited by Henry Harris, pp. 13–45. Oxford: Oxford University Press.

Ramaiah, C. 1978. *The Problem of Change and Identity in Indian Philosophy.* Tirupati: Śri Venkateswara University Press.

Ruegg, D. Seyfort. 1977. "The Uses of the Four Positions of the Catuṣkoṭi and the Problem of the Description of Reality in Mahāyāna Buddhism." *Journal of Indian Philosophy* 5 : 1–71.

Shoemaker, Sydney, and Richard Swinburne. 1984. *Personal Identity.* Oxford: Basil Blackwell.

Siderits, Mark. 1987. "Beyond Compatibilism: A Buddhist Approach to Freedom and Determinism." *American Philosophical Quarterly* 24 : 149–159.

Stone, Jim. 1988. "Parfit and the Buddha: Why There Are No People." *Philosophy and Phenomenological Research* 48 : 519–532.

Unger, Peter. 1979. "Why There Are No People." *Midwest Studies in Philosophy* 4, edited by Peter French, Theodore Uehling, and Howard Wettstein (Minneapolis: University of Minnesota Press): 177–222.

Vijñaptimātratāsiddhi of Vasubandhu. 1980. Edited and translated by K. N. Chatterjee. Varanasi: Kishor Vidya Niketan.

Visuddhimagga of Buddhaghosha. 1950. Edited by Kosambi and Warren. Harvard Oriental Series, vol. 41. Cambridge: Harvard University Press. Translated as *The Path of Purification* by Bhikkhu Ñāṇamoli. Berkeley: Shambala, 1976.

John A. Taber

The Mīmāṃsā theory of self-recognition

The defense of the self, that is, of a continuously existing individual self through time, is one of the more intriguing features of the Mīmāṃsā system. Besides posing formidable problems for the Buddhist opponents of the self theory, it is of direct relevance to the ongoing Western debate about personal identity. Śabarasvāmin (fifth century) and Kumārilabhaṭṭa (seventh century), the two Mīmāṃsā thinkers I shall treat here, develop a position which contemporary Western theorists of personal identity tend to overlook. Yet it is, at least, not patently untenable, and it appears to have been held by some of the earlier philosophers who treated personal identity. In this article I shall expound the Mīmāṃsā position—or, more exactly, one particular argument of that position which I believe to be of special interest—and then relate it to the Western discussion.

My purpose in this comparative study is twofold. First, I seek to understand the doctrine of the self as presented in the original Mīmāṃsā texts. To do so, one must achieve a *philosophical* understanding of the doctrine; one cannot interpret a text adequately—be it philosophical or medical or mathematical, and so forth—unless one understands in some measure independently of the text what it is about. Comparison of the presentation of a philosophical idea in one text with presentations of the same or similar ideas in other texts stimulates philosophical understanding of it, insofar as to comprehend something is to know to what it is similar and from what it is dissimilar. Cross-cultural comparisons often yield more varied contexts in which an idea, in various guises, is discussed, revealing more of its ramifications—as well as, of course, relating it to the familiar. As Bhartṛhari notes, "insight attains clarity through the study of diverse traditional views" (*prajñā vivekaṃ labhate bhinnair āgamadarśanaiḥ—Vākyapadīya* 2.484). All this goes without saying for most of the readers of this journal, but there is no harm in repeating it. My second, subsidiary, purpose is to recover a lost argument from the history of Western philosophy. The study of another philosophical tradition often affords a philosopher new perspectives on types of reasoning that exist in his own but, for whatever reasons, have been forgotten or left undeveloped. The discovery of a more rigorous, or just somewhat different, formulation of an argument in another tradition can revive interest in it. Salvaged and refurbished, it can sometimes be made use of in contemporary discussion. Precisely such an argument, I believe, is that for the existence of the self, which I discuss in this article. However, I do not attempt to apply it in actually solving any aspect of the modern debate about personal identity. I am content here merely to identify it as prima facie relevant.

John A. Taber is Assistant Professor of Philosophy at the University of New Mexico.

AUTHOR'S NOTE: *A version of this article was read at the Sarvapelli Radhakrishnan Centennial Conference, Oxford, Ohio, April 1988.*

I

The Mīmāṃsā argument in question is developed from a certain Nyāya argument for the existence of the self. Śabara, in his discussion of the self in his commentary on *Mīmāṃsā Sūtra* 1.1.5, considers several Nyāya arguments and their criticisms. The last of these is none other than the argument expressed in *Nyāya Sūtra* (*NS*) 1.1.10: *icchādveṣaprayatnasukhajñānāny ātmano liṅgam iti.* This *sūtra* is by no means easy to interpret—the commentator Uddyotakara himself offers three ways[1]—but one reading of it, the one adopted by Śabara, is as follows: The existence of a continuously existing, substantial self can be inferred from the occurrence of desire, aversion, effort, and so forth, insofar as these states are directed towards objects that have been experienced to be pleasant or unpleasant in the past. Now, the Naiyāyika assumes, one certainly cannot desire what someone else has experienced in the past. Therefore, in order for desire and so forth to occur, there must be one self existing continuously from the past to the present who both had the pleasurable experience of the object and desires it now. As Śabara and others who have discussed this argument make clear, it ultimately rests on the claim that one cannot desire a thing one does not *remember* (as yielding pleasure). And one cannot remember *someone else's* experiences. For memory to occur there must be a single, continuously existing substance who both had the remembered experience and remembers it.[2] I shall, therefore, refer to this argument as the argument from memory.[3]

What is wrong with this argument? Why is the Mīmāṃsaka not content with it? Its weakness is just the premise that a memory of a past experience is possible only if there is a continuously existing substance that both had the experience and is now remembering it. For the Buddhists developed a plausible alternative account of memory. A person might not be a single substance existing continuously through time but a series of what, borrowing from contemporary philosophy, we might call person stages. Each person stage, consisting of various psychological and physiological factors (*skandhas*), lasts for only an instant, but as it passes away it gives rise to another person stage having important continuities with it. (A "person stage" as usually defined—a phase in the series of states and events which we ordinarily call a "person"—does not presuppose the existence of a continuous substance.) Memory, in this theory, would be the result of a chain of memory impressions (*vāsanās*) occurring in a series of person stages, initiated by an original experience occurring at person stage *A* and eventually giving rise to a memory experience (an experience similar to the original experience or having it as its object) at a subsequent person stage *Z*. Thus, the Nyāya argument from memory is not conclusive (*anekānta*). Memory experiences might take place without a continuous, substantial self.[4]

The Mīmāṃsaka considers this criticism as valid. Kumārilabhaṭṭa main-

tains that Śabara, in discussing the various Nyāya-Vaiśeṣika arguments for the self, merely intends, by means of Buddhist objections, to point up their inadequacies.[5] He moves beyond all of them to adopt a uniquely Mīmāṃsā position, which the Buddhist is unable to assail. Here, as usual, the Mīmāṃsaka is asserting himself to be the only competent defender of the faith among orthodox Hindu thinkers. The irony here, as elsewhere (in the discussion of God's existence, the defense of the authority of the Veda, and so forth), is that he nearly ends up agreeing on more points with the Buddhist than with the other representatives of the Brahmanic tradition.

What, then is Śabara's argument? It is but a revision of this Nyāya argument from memory. That is, Śabara argues that a substantial self is made evident to us by the very notion 'I' (*ahaṃpratyaya*) as it occurs in memory. I quote now from the *Mīmāṃsāsūtrabhāṣya*:

> When a thing has been seen on one day, the notion "*I* saw this" arises on a later day. And this [recognition] refers to an inner self (*pratyagātmani*), nothing else [that is, certainly, not to a collection of *skandhas*]. For the present [collection of *skandhas*] is different from the one that saw [the thing] earlier. Hence, there is something besides the [collection of *skandhas*] to which this word 'I' applies. . . . [Here,] we do not consider the *word* 'I' that we employ to be the means for inferring another thing [besides the *skandhas*]. Rather, we consider the *experience* of recognition (*pratyabhijñā*), which goes beyond the word (*śabdād vyatiriktam*) [to be the means of inference]. [So, one could not argue that no conclusion can be drawn from the use of 'I' here because it is used only figuratively.] For this is the meaning we comprehend [when we employ this word]: "We perceived yesterday, we remember now." Therefore, we understand that we existed yesterday and we exist today. And those things which existed yesterday and exist today, they are not such as have passed away. [That is, they exist *continuously* from yesterday to today.][6]

The thrust of the argument is this: Memory establishes the existence of a continuous self not insofar as it *presupposes* a single subject of experience that both had the remembered experience and now remembers it, but insofar as it directly reveals one. For it is part of the content of many of my memories that *I*, who am now remembering, am the one who did or experienced the thing that is remembered. And that means that it is part of the content of many of my memories that I, who am remembering now, am identical with a thing—a subject of experience—that existed in the past. From this I infer that I have existed continuously through time. (Thus, technically, we still have to do with an inference. The ground of inference (*hetu*) is the recognition of myself as identical with a thing that existed in the past, the conclusion of my continuous existence through time.) Just as the establishing of any other object that exists now that it existed also in the past justifies the belief that it has existed continuously from the past to the present, so does the *recognition* of myself as a subject of experience in memory. We may refer to this as the argument from self-recognition.[7]

Kumārila adds very little to this. The *ātmavāda-adhikaraṇa* of his *Śloka-*

vārttika comprises only 148 *ślokas*, even though Śabara devotes considerable space to his discussion of the self. Kumārila may have thought the self a more appropriate topic for Vedānta than Pūrva Mīmāṃsā (thus, see the concluding *śloka* 148). Most of his effort goes to establishing that the thing that is identified as 'I' is not the physical organism, including the mind and the senses. For the 'I' is an agent of knowledge (*jñātṛ*)—it is one who has *experienced* something in the past—and the physical organism is incapable of knowledge by itself. Nor can this recognized entity be a cognitive state (*jñāna*), as a (Yogācāra) Buddhist might argue. For cognitive states are fleeting. The same cognitive state that 'experienced' something in the past cannot be 'remembering' now. But one need not have gone to any length arguing for these things. They are immediately conveyed by Śabara's formula, if interpreted literally: the notion 'I' in memory refers to (or, "arises with respect to") an inner self and nothing else (*pratyagātmani caitad bhavati, na paratra*). The idea seems to be that I recognize myself as a distinct kind of entity, not as a body nor even as a subject of experience in general, for which the expression 'I' is required. The argument may, then, be seen to depend partially on another Nyāya view, namely, that the word 'I', insofar as it is a unique word different from the words 'body', 'idea', and so forth, requires a unique occasion for its use.[8] This is Hume's argument against the existence of a substantial self turned on its head: there *must* be some distinct impression corresponding to the notion 'I'; otherwise the notion would not occur. Moreover, Śabara adds, any collection of *skandhas*—including physical form, that is, the body and sense organs—changes over time. But in memory I recognize the same subject as my present self—I recognize *me*—not something just similar or related to me. That is, I recognize not just a distinct type of entity but a distinct particular.

From what I have said, the Mīmāṃsā argument for the self might be interpreted as a merely linguistic one: from the way we talk about things we infer the way things are. This is a notoriously unreliable kind of argument. Certainly, the way we talk does in some measure reflect our experience of the world, but not always nor in all respects. But I believe that it would be a mistake to interpret the argument in this way. As I see it, in spite of its form as an inference, it amounts to the assertion that our way of talking about the self, specifically, the reference to ourselves in our memory reports as continuously existing and self-identical, is indeed an accurate expression of our experience.[9] Viewed in this way, the argument comes down to a direct appeal to a certain intuition which, if acknowledged, ought to settle the issue immediately. If one does not acknowledge it, on the other hand, then the dispute about the self is exposed as unresolvable. I, for one, am inclined to agree with the Mīmāṃsā analysis of memory. It at least seems truer to my experience than those accounts which suggest that when I remember something, I recall *some subject of experience or other* having undergone the experience.

Rather, it seems to me that when I remember something, I do recognize *me* undergoing it. And that would seem to exclude my being a body or a cognitive state, which could under no circumstances be quite identical with the present body or cognitive state (especially when I am remembering something that took place long ago). In any case, in recognizing myself I do not, evidently, recognize a certain body or an idea. The Mīmāṃsā account also entails that whenever I have an experience, I am aware of myself having it. Otherwise, I could not later, in self-recognition, identify the subject of the experience as myself. Thus, the Mīmāṃsā analysis of memory experience—the claim that it entails a self-recognition—is supported in some measure by the widespread belief that when we know something, we know that we know it.

So far, in discussing the Mīmāṃsā argument from self-recognition I have, while pointing out various ways in which it relates to Nyāya ideas, gone along with the Mīmāṃsā claim that this argument constitutes an advance beyond Nyāya thought about the self. But is that really true? In the first *āhnika* of the third *adhyāya* of the *Nyāya Sūtra* and its commentaries, we in fact find a discussion that comes very close to expressing the argument from self-recognition.[10] I now turn to this discussion with the purpose of futher clarifying the Mīmāṃsā point of view.

NS 3.1.1 reads: *darśanasparśanābhyām ekārthagrahaṇāt.* Vātsyāyana and Uddyotakara interpret this as an answer to the question whether, from statements such as "I see with my eyes," "I know with my mind," and "I feel pleasure and pain with my body," one can infer that there is a self that is distinct from mind, body, and senses. For such statements suggest that there is an agent that employs the senses, the mind, and the body in cognitive acts, and usually agent and instrument are distinct entities. The commentators explain, however, that it is possible that such statements merely express a relationship between part and whole; that is, they may be referring to a relationship, not between a particular faculty and some altogether different entity (a self), but between a particular faculty and the collection of all the faculties that constitute the physical organism (*dehādisaṅghāta*). (On this view we say, "I see with my eyes," just as we might say, "The house stands by means of its pillars.") The *sūtra*, then, responds to this possibility by asserting that there must be a distinct entity which employs these faculties as instruments in acts of cognition, "because one and the same object is apprehended by both [faculties of] seeing and feeling." Vātsyāyana elucidates: "The very same object which was [formerly] apprehended by the sense of sight is [now] apprehended by the sense of touch. [Thus, we express our experience by saying,] 'What I saw with my eye I now also feel with the sense of touch, and what I felt with the sense of touch I now see with my eye.' These two cognitions [of seeing and touching] are joined together (*pratisandhīyete*) as having the same object and the same agent; and they do not have an aggregate as their agent, nor is the sense organ [itself] the agent." The point is: there must be some one entity

distinct from any sense organ which both sees and feels the object. For one sense organ cannot perceive the object of another: an eye, for example, cannot feel. Nor could the *aggregate* of faculties be responsible for these acts of cognition, for it is *one* thing, not a group, that perceives the object on separate occasions. And, the aggregate consisting solely of various faculties, one would still have the absurdity of one faculty cognizing the object of another. Thus, Vātsyāyana concludes, "That [thing] which is the perceiver of the same object by means of [both] the eye and the sense of touch, which combines together the two cognitions which have the same object and the same agent but different efficient causes [that is, which are produced by different sense organs]—that is the self."[11]

Now this argument seems indeed very closely related to the Mīmāṃsā argument from self-recognition. It appeals to the same basic fact on which the latter turns, namely, that the same subject of experience who cognized something in the past may cognize that thing now. At *NS* 3.1.7 the term 'recognition' even occurs: *savyadṛṣṭasyetareṇa pratyabhijñānāt*. Vātsyāyana interprets this to mean that we know that there is a bearer of consciousness distinct from the body and the various faculties because we recognize with the right eye what we once saw with the left! If the right or left eye were the bearer of consciousness, this could not occur, for, once again, one faculty cannot recognize what has been experienced by another (the two eyes are here, oddly enough, being considered as distinct faculties).[12] Nevertheless, this argument does not appeal to the more specific fact on which the Mīmāṃsā discussion focuses, that the *self* is recognized. Self-recognition directly reveals the existence of the self in the past and the present. But in all the examples given of memory or recognition by Vātsyāyana it is objects of sense that are recognized or remembered, while the self is still being *inferred*, essentially as in the discussion of *NS* 1.1.10, as the single entity continuous from past to present, which recognition and memory must presuppose. Moreover, the ultimate aim of *NS* 3.1.1 is only to prove that the self—whose existence as the continuous *pratisandhātṛ* (connector) of memories was, in fact, supposedly already established at 1.1.10—is an entity *distinct* from the aggregate of body and faculties. Thus, Vātsyāyana stresses examples that involve experiences of various kinds—for example, "What I *saw* in the past I am *touching* now"—where it is clear that a particular sense organ could not be the subject of the experiences. But for the Mīmāṃsaka the argument from self-recognition establishes only the *permanence* of the self. The explicit demonstration of a distinction of the self from the faculties is then carried out, it seems, as an afterthought, by a different method than that of Nyāya, namely, by appeal to the notion that a material organism and its faculties cannot be bearers of consciousness.

At *NS* 3.1.14, however, Vātsyāyana actually refers to the fact that every memory contains not just an awareness of an object previously experienced,

but an awareness of the object *as* previously cognized by an agent of cognition: "This memory which occurs with reference to an object not presently apprehended, in the form 'I knew that thing'—the object of this memory is a previously known thing, qualified [as being known] by an agent of cognition [through] a process of cognition (*jñātṛjñānaviśiṣṭaḥ pūrvajñāto 'rthaḥ viṣayaḥ*), not just the thing."[13] Here, again, the context is the Buddhist theory of memory. The Buddhist has charged in the previous *sūtra* that the fact that a perception of an object by means of one sense can stimulate a memory relating to another sense—for example, that seeing a mango can cause one's mouth to water—cannot be adduced in support of the existence of a self which is distinct from the sense faculties (the argument would be basically the same as that of 3.1.1 and 7: there has to be some third entity capable of perceiving the objects of both faculties); for the self, the Buddhist argues, is not part of the content of memory—only the sense object, the mango, is (*NS* 1.1.13: *na, smṛteḥ smartavyaviṣayatvāt*). Thus, it is reasonable to think that only the mango, experienced in a certain way in the past and now presently being perceived by some other sense, causes the memory; we have no evidence that a self has a role to play. Vātsyāyana then interprets *NS* 3.1.14 (*tadātmaguṇasadbhāvād apratiṣedhaḥ*, "Because of the existence [of memory] as a quality of the self, one cannot deny [the self]") as yet another expression of the idea that memory is possible only if it belongs to a continuous self, for one thing (faculty) cannot remember what another has experienced. But Vātsyāyana goes on to say that the Buddhist is also wrong in thinking that the object of memory is merely the thing that was previously experienced. Memories of an absent object in fact apprehend the object *as having been experienced in an act of knowledge by a knowing subject.* Thus, typically, we say something like, "I knew that thing," or, "That thing was known by me."[14] In memories with respect to an object at hand (*pratyakṣe 'rthe yā smṛtiḥ . . .*)—that is, in acts of recognition—more than one cognition pertaining to a single object are expressly attributed to a single subject: "I am seeing the thing I saw before." Thus, Vātsyāyana suggests, in memory a single self comes to the fore as a factor in several experiences.

Here, it seems, the Nyāya philosopher Vātsyāyana has gotten even closer to the fact that is central to the Mīmāṃsā discussion of the self, namely, that in recognition we ascribe various experiences to one and the same self; the past existence of a subject of experience is part of the content of memory. But he still stops short of saying that the *self* is *recognized* in memory. The object of recognition here is still some other thing. (Recall, on the other hand, how Śabara formulates the fact of recognition: "*We* perceived yesterday, *we* remember now.") And Vātsyāyana is still appealing to our ascribing various experiences to a single self only as a basis for an inference to a self distinct from the body and the faculties: a particular *faculty* could not be the subject

of various kinds of experiences; therefore it must be some other kind of thing. One way to appreciate the difference between the Mīmāṃsā and Nyāya approaches is to note that it can be asked precisely at this point of the Naiyāyika—as in fact the Buddhists did ask (and, as we shall see below, as Locke asked in his day): Granted that we do *ascribe* several experiences to one subject, how do we *know* that we have to do with the same subject in every case? How do we know that the entity we repeatedly experience as 'I' is the same in every cognition? For certainly it is conceivable that in different cognitions one has to do with qualitatively indistinguishable but numerically different things. The Naiyāyika has no answer to this question, but the Mīmāṃsaka answers it, in effect, by asserting that, indeed, we know that we have always to do with the same entity because we *recognize* it from cognition to cognition. Thus, the proof of the permanence of the self—that from moment to moment there exists only one subject of experience—is placed on a firmer footing.[15]

Note also that the basic Nyāya argument presented by *NS* 3.1.1 for proving the distinction of the self from the faculties is fallacious. It may well be that one sense organ cannot apprehend the object of another, but that does not rule out that a collection of various faculties could. In the case of an army, while it is certainly true that the artillery cannot charge and the cavalry cannot bombard, it is nevertheless the case that the army can both charge with its cavalry and bombard with its artillery; and no one believes that an army consists of anything more than its artillery, cavalry, infantry, and so forth. So, similarly, there seems no reason why a mere aggregate of faculties, including a mind and various sense organs, could not think with its mind, perceive one kind of thing with one sense and another kind of thing with another, and yet be the *same* thing which performs each act. Thus, this kind of argument is inconclusive and fails to secure the distinctness of the self. Here, too, the Mīmāṃsā argument seems more fit for success insofar as it entails that, in self-recognition, we cognize the self directly as a single thing that remains identical over time. That rules out all other candidates which could undergo any measure of change, such as the body, mind, and sense faculties.

In the end, however, we must be aware that the Nyāya and Mīmāṃsā arguments are very closely related. If the Mīmāṃsaka 'discovered' the argument from self-recognition, he did so only by looking at the same facts to which the Naiyāyika had already called attention and giving them a slightly different twist—or perhaps, indeed, the Naiyāyika was originally introduced to those facts by the Mīmāṃsaka. In general, a comparison of these arguments demonstrates how Nyāya and Mīmāṃsā, at this stage in the history of Indian philosophy, developed in close interaction with each other.

It remains, before we move on to discuss Western parallels to the doctrine of self-recognition, to consider briefly the Buddhist response to the Mīmāṃsā

position. This is to be found in Śāntarakṣita's *Tattvasaṅgraha* (*TS* 241–284). Śāntarakṣita, quoting Kumārila at length (some of the *kārikās* attributed to the Mīmāṃsā but not locatable or precisely matched in Kumārila's *Ślokavārttika* may well be from his longer, lost commentary on the *Śābarabhāṣya*, the *Bṛhaṭṭīkā*), directs most of his criticism against the Mīmāṃsā belief that the self is an eternal, essentially conscious entity that continues (essentially) unchanged throughout its various transformations (states of consciousness), as a snake remains the same in passing from a coiled to an uncoiled state (*TS* 223–225). Śāntarakṣita asks, if consciousness is truly single and eternal, then how can a multitude of fleeting ideas appear within it (*TS* 241)? Moreover, if pleasure, pain, and so forth, as states of consciousness, are not altogether distinct from the self, then when they arise and disappear so must the self arise and disappear (*TS* 268). In short, the Buddhist, here as elsewhere, is unable to conceive of continuity through change; indeed, he argues that such a possibility is incoherent. "The snake becomes curved, etc., because it is subject to perpetual flux; if it had a permanent form, then, like the soul, it could never come by another state" (*TS* 274, Jha's translation, amended). With regard to the argument from self-recognition Śāntarakṣita argues in general—initially against the Naiyāyika—that neither the word 'self' (*ātman*) nor the word 'I' (*aham*) is a referring expression. 'Self' is only a conventional term for mind or consciousness (*citta*), which is a fleeting factor of the life-stream according to the Buddhist. It has no unique referent. Verbal usage being established merely by convention, words are often used to talk about nonexistent things (*TS* 204–206). Similarly, the word 'I' does not report a self-*perception*. No such thing as a self—having eternal consciousness, omnipresence, and so forth—is manifest to us when we employ that word. Were that the case, there would be no dispute about its existence (*TS* 212–216). These points then apply directly to the argument from self-recognition. Nothing is recognized when we think, "I am the one who experienced this in the past." For the notion 'I' is just an idea without any basis, originating from the beginningless tendency to ascribe (erroneously) substantive being to things (*TS* 275–284).[16]

It should be noted that the doctrine of the self-perceived character of cognitions (*svasaṃvedyatva*), crucial to Yogācāra epistemology, is not brought into this discussion. It is not appealed to as, say, the basis of an alternative analysis of self-consciousness. Although Kumārila, as we saw, saw fit to deny that 'self-recognition' could be the cognizing of one (earlier) idea by another (later) idea, Śāntarakṣita does not take up the gauntlet over this. Rather, he simply relegates the *self*-consciousness that the Hindu philosopher has in mind—the awareness of a unique type of thing called 'self'—to the realm of delusion. Further below we shall have occasion to consider the Mīmāṃsā response to this maneuver.

II

I shall now attempt to show how the sort of view developed by the Mīmāṃsā philosophers—the doctrine of self-recognition in memory—fits into the Western discussion of personal identity.

The approach to the self in the West has been somewhat different from that of Indian philosophy. Indian thinkers have been primarily concerned with the soul or self, that is, whether there is a continuous, spiritual substance distinct from body, mind, and senses to which the word 'self' refers. Western philosophers, on the other hand, have been primarily concerned with the common concept of a 'person', specifically, with the identity of what we refer to as a 'person' through time.[17] The main question in the West has been: What is the (necessary and sufficient) criterion (or criteria) that determines two person stages to be stages of one person?[18] In recent times it has generally been held that this question can be answered without going into the problem of whether a person is a single, continuous substance; for most Western philosophers have taken for granted the (Buddhist) possibility that whatever relation between person stages makes them stages of one person could hold between entities that are distinct as substances. Nevertheless, at the beginning of the Western discussion of personal identity there was considerable debate about this issue, and it is this aspect of the discussion—which of course relates more directly to the Hindu-Buddhist debate about the self—on which I shall focus. We shall see that it indeed has bearing on the more recent discussion of personal identity.

John Locke is generally credited with being the first to have proposed the memory criterion of personal identity. According to this criterion two person stages are of the same person if one contains memories of experiences that belong to the other. (In recent times this position has been developed by Paul Grice and Anthony Quinton.)[19] But, though most of what Locke writes is consistent with this view, it is important to see specifically why he held it. It does not appear that he held it because he believed in a version of the Nyāya argument from memory, that a memory and the past experience which is remembered require a single, continuous substratum from past to present which has both. Rather, Locke seems to have held the view that what makes an experience as of a certain person is that it is undergone with a consciousness of oneself as that person, and in cases of memory the present remembering and the past experience which is remembered are both experiences that, insofar as they are attributed to the same self, are undergone with the same awareness of self. "Wherever a man finds himself," Locke writes, "there, I think, another may say is the same person."[20] On the other hand,

in like manner, it will be in reference to any immaterial substance, which is void of that consciousness whereby I am myself to myself: if there be any part of its existence which I cannot upon recollection join with that present con-

sciousness, whereby I am now myself, it is in that part of its existence no more myself, than any other immaterial being. For whatsoever any substance has thought or done, which I cannot recollect, and by my consciousness make my own thought and action, it will no more belong to me . . . than if it had been thought or done by any other immaterial being anywhere existing.[21]

Thus, for Locke, it is not just that an experience of a previous person stage is *remembered* at the present stage that determines both to be stages of one person, but that the experience of the previous person stage is remembered as having been accompanied by the same self-consciousness as accompanies the present person stage.[22] This should be kept in mind whenever the following passage, often cited as evidence that Locke subscribed to the memory criterion, is read:

. . . As far as any intelligent being can repeat the idea of any past action with the same consciousness it has of any present action; so far is it the same personal self. For it is by the consciousness it has of its present thoughts and actions, that it is self to itself now, and so will be the same self, as far as the same consciousness can extend to actions past or to come. . . .[23]

Read consistently with his other statements, "the same consciousness" to which Locke refers is the same *self*-consciousness, the consciousness of the experience as being *mine*.

Thus, it would appear that Locke is putting forward a verson of the doctrine of self-recognition in memory. But it is no more that than are the Nyāya arguments which take note of the fact that we typically ascribe several experiences, past and present, to one subject. The recurrence of self-awareness Locke has in mind is not a self-recognition. Rather, it seems that for Locke two person stages are of one self only insofar as they are each accompanied by a consciousness which *imputes* them to one and the same subject—a feature of those experiences that need not correspond to any objective, physical or metaphysical, fact. Thus Locke, notoriously, was able to consider that this consciousness of self could be transferred from one substance to another. That two person stages are undergone with the same consciousness of oneself does not prove that they are of the same identical substance: ". . . It being the same consciousness that makes a man be himself to himself, personal identity depends on that only, whether it be annexed solely to one individual substance, or can be continued in a succession of several substances."[24]

It was especially with regard to this point that Butler and Reid criticized Locke. They charged that he misconstrued the recurrence of "the same [self-] consciousness." It is not just a matter of being aware of two experiences as occuring together with the same self-consciousness or as ascribed to one self, but of *recognizing* the subject of the remembered experience as the present rememberer, as one would recognize any other object. Hence, there can be no question that the person is a thing that continues from past to present—

that is, is a single substance. Joseph Butler attributes Locke's mistake to the fact that the *acts of self-consciousness* in different experiences are numerically distinct:

The ground of doubt, whether the same person be the same substance, is said to be this; that the consciousness of our own existence in youth and in old age, or in any two joint successive moments, is not *the same individual action*, i.e., not the same consciousness, but different successive consciousnesses. Now it is strange that this should have occasioned such perplexities. For it is surely conceivable, that a person may have a capacity of knowing some object or other to be the same now, which it was when he contemplated it formerly; yet in this case, where. . . the object is perceived to be the same, the same perception of it in any two moments cannot be one and the same perception.

And that is precisely what we have in the case of memory— two perceptions, one in the past and one now, of one self as the subject of certain experiences then and now:

. . . Though the successive consciousnesses which we have of our own existence are not the same, yet they are consciousnesses of one and the same thing or object; of the same person, self, or living agent. The person, of whose existence the consciousness is felt now, and was felt an hour or year ago, is discerned to be, not two persons, but one and the same person; and therefore is one and the same person.[25]

Here, at last, we really do have the doctrine of self-recognition in a philosopher other than a Mīmāṃsaka. Even more explicit is the presentation of Thomas Reid, who claims only to be following Butler. "How do you know," Reid asks, "that there is such a permanent self which has a claim to all the thoughts, actions, and feelings, which you call yours?" He replies:

. . . The proper evidence I have of all this is remembrance. I remember that twenty years ago I conversed with such a person; I remember several things that passed in that conversation; my memory testifies not only that this was done, but that it was done by me who now remembers it. If it was done by me, I must have existed at the time, and continued to exist from that time to the present. . . . Every man in his senses believes what he distinctly remembers, and every thing he remembers convinces him that he existed at the time he remembered it.[26]

We need not be misled by Reid's choice of the word 'remembrance' instead of 'recognition'. To identify a thing as a thing that one is already acquainted with is, technically, in Indian philosophy, to recognize (*praty abhi* $\sqrt{\textit{jñā}}$) that thing, and that is what Reid is talking about (identifying the subject of a remembered experience as the present remembering self). His statement here is remarkably close to that of Śabara quoted in the first part of this article. As a final piece of evidence that the doctrine of recognition was held by these philosophers, I present another passsage from Butler where he, too, explains how one arrives at the permanence of the self from the fact of self-recognition:

> Every person is conscious, that he is now the same person or self he was, as far back as his remembrance reaches; since, when any one reflects upon a past action of his own, he is just as certain of the person who did that action, namely himself, the person who now reflects upon it, as he is certain that the action was at all done. . . . And this he, person, or self, must either be a substance, or the property of some substance. If he, if person, be a substance; then consciousness that he is the same person, is consciousness that he is the same substance. If the person, or he, be the property of a substance; still consciousness that he is the same property, is as certain a proof that his substance remains the same, as consciousness that he remains the same substance would be; since the same property cannot be transferred from one substance to another.[27]

Both Butler and Reid emphasize that my remembering that I did something does not *make* me the one who did it (a view they attribute to Locke); rather, I am able to remember it because I was indeed the subject that did it in the past. Self-recognition does not constitute self-identity objectively; it is only the criterion for *ascertaining* whether two person stages are of one person.[28] Butler and Reid also stress that our awareness, by this means, of our identity with our past selves is more complete than for other objects. We recognize ourselves, as subjects of past experiences, always as precisely identical to ourselves now, not more or less identical as, say, a tree is to the sprout from which it grew. Reid attributes this to a person being a "monad," not divisible into parts, whereas other things change by parts being added, subtracted, or replaced. Hence, they suggest that the general concept of identity is actually based on the more specific notion of self-identity, as true identity does not pertain to other kinds of things.

In the passage from Reid cited above we find him asserting that "every man in his senses believes what he distinctly remembers." Here, another parallel to the Mīmāṃsā view emerges. Reid's statement relates to an obvious objection to the doctrine of self-recognition: Granted that one has a sense of self-recognition when one remembers oneself doing something in the past, how does one know that one has not made a mistake, that what one takes to be precisely oneself (then) is but some other entity closely resembling oneself (now)? (that is, one could have to do with two qualitatively indistinguishable entities which are able to cause, one in the other, states which give rise to memory experiences). This is similar to the question: How does one know that the bird one sees now sitting on a branch is the same bird one saw there a moment ago, not another bird with identical markings? Reid's answer to this, based on his general discussion of memory, is that we are generally incapable of disbelieving what we distinctly remember, so that in an actual case, where one remembers that it was *oneself*, not someone else, who did such and such, the skeptical doubt simply does not come up. And that is as much as to say that our whole system of knowledge is based on accepting what memory presents to us definitively, without asking for futher evidence—otherwise we would get caught up in a regress. Butler puts the matter quite cogently:

Every person is conscious, that he is now the same person or self he was, as far back as his remembrance reaches; since, when any one reflects upon a past action of his own, he is just as certain of the person who did that action, namely himself, the person who now reflects upon it, as he is certain that the action was at all done. . . . But though we are thus certain that we are the same agents, living beings, or substances, now, which we were as far back as our remembrance reaches; yet it is asked, whether we may not possibly be deceived in it? And this question may be asked at the end of any demonstration whatever; because it is a question concerning the truth of perception by memory. And he who can doubt, whether perception by memory can in this case be depended upon, may doubt also, whether perception by deduction and reasoning, which also include memory, or, indeed, whether intuitive perception can. Here then we can go no further. For it is ridiculous to attempt to prove the truth of those perceptions, whose truth we can not otherwise prove, than by other perceptions of exactly the same kind with them, and which there is just the same ground to suspect; or to attempt to prove the truth of our faculties, which can not otherwise be proved, than by the use or means of those very suspected faculties by themselves.[29]

All this is parallel to Kumārila's discussion. Anticipating the Buddhist objection that the notion 'I' in self-recognition is without any basis, Kumārila asserts—rather dubiously—that it cannot be erroneous insofar as it is produced by a memory impression (*vāsanā*); for a memory impression always gives rise to an idea faithful to the experience that caused it (*ŚV* 124–125).[30] But, more cogently, he goes on to point out that the notion 'I' (in general, or specifically as it occurs in self-recognition?) is not overturned by a subsequent cognition (*ŚV* 125–126). (That is, applied specifically to self-recognition, one does not, subsequent to believing that one has recognized oneself as the subject of a previous experience, come to believe that the subject of that experience was not oneself but, say, merely something similar to oneself.) According to Mīmāṃsā epistemology, to consider a cognition invalid one must realize either that it does not present things as they really are, by having another cognition that contradicts it, or else that the faculty of knowledge that gave rise to it is defective. In short, the position of Butler and Reid harmonizes with the Mīmāṃsā theory of the intrinsic validity (*svataḥ prāmāṇya*) of cognitions. According to that theory, every cognition presents itself to us initially as true without requiring the evidential support of another cognition. Were the application of another means of knowledge required to establish the truth of any cognition, one would have a regress: the confirming cognition itself would require confirmation, and so on ad infinitum.[31] This is quite similar to the basic principle of Reid's general theory of knowledge, given expression by Butler in the passage above, that a belief grounded in perception (which for him includes memory) is innocent until proven guilty.[32]

The ideas of Butler and Reid about the self have not been completely lost in Western philosophy. Various contemporary philosophers have made use of them in various ways in the discussion of personal identity. Thus, Sydney Shoemaker has argued—explicitly acknowledging a debt to Butler and

Reid—that our identifications of ourselves as agents of past actions or the subjects of past experiences are "non-criterial." In saying, for example, "I broke the front window yesterday," one does not employ criteria in identifying oneself as the one who broke the window. If criteria were employed—if one consulted certain facts in a process of judgment leading to the conclusion that it was, indeed, oneself who broke the window—then it would also be possible *not* to come to that conclusion. And that, especially in a case where no one else is remembered as being present, would be absurd. Rather, Shoemaker suggests, one knows directly that it was oneself who broke the window without appeal to any criteria; for one remembers *oneself* breaking the window. With regard to another example, Shoemaker writes,

> The sentence "I remember having a headache yesterday" [which, again, might suggest to some that I can first remember an experience and then meaningfully ask *who* underwent it, which question would then be answered by consulting criteria] does not differ in meaning from the sentence "I remember my having a headache yesterday." But if what I remember when I remember a past headache is *my having* a headache, or that *I* had a headache, my statement "I had a headache" is [just] a memory statement, not a conclusion *from* what I remember, and cannot be grounded on any criterion of identity.[33]

Shoemaker also specifically denies that one adduces the fact that *one remembers* the experience in question as a basis for judging that it was undergone by oneself, on the principle that one can only remember what oneself has experienced. For that one remembers a past experience is not, from the first-person point of view, something contingent—a possible distinguishing feature of an experience which one can determine to belong to the experience by scrutinizing it. If that were the case, it would also be possible to determine of a past experience that one does *not* remember it, which would be absurd! So there is no room for employment of the fact that one remembers an experience as a criterion that it was undergone by oneself. Rather, one simply remembers oneself undergoing it. Thus, Shoemaker rejects the "memory criterion" of personal identity, as originated by Locke and revised by Grice: the fact that one person stage contains a memory of an experience occuring in another person stage is not something we appeal to as a *criterion* of their being the same person.[34]

Nevertheless, Shoemaker's position, in the end, is a far cry from the doctrine of self-recognition. For he denies that we ever have direct self-acquaintance or self-knowledge, which is clearly entailed by that doctrine. To *recognize* oneself one must be directly acquainted with oneself now and have been in the past.[35] Shoemaker believes, rather—influenced here by Wittgenstein—that we make statements about ourselves—ascribe perception and memory experiences to ourselves—noncriterially simply as a result of having been trained to make certain utterances in certain circumstances.[36] Thus, conversely, the doctrine of self-recognition represents a certain criti-

cism of Shoemaker's position. If we do actually recognize ourselves in memory, then our use of 'I' is, although non-criterial, still in some way the expression of an act of self-apprehension.

More faithful to Butler and Reid is Geoffrey Madell, who in a recent book has called for the "vindication of the Reid/Butler view of personal identity."[37] Among other things Madell, building on important work by Hector-Neri Castañeda, argues that the word 'I' is a true referring expression—clearly a presupposition of any theory of self-recognition.[38] Madell, too, however, ultimately diverges significantly from Butler and Reid. He does not pick up on their self-recognition thesis; in any case, it appears that he would not go along with it. Although he agrees that "personal identity through time is unanalysable; [that] what unites my experiences [makes them as of one person], whether they are simultaneous or spread over time, is just their being mine," he claims that this does not entail that there is a substantial ego.[39] But, as we saw Butler argue above, if there is self-recognition, then there is an *enduring* (substantial) self.

Perhaps closest to the ideas of Butler and Reid are the views of Roderick Chisholm. In his book *Person and Object*, Chisholm goes against Shoemaker in claiming that the use of the word 'I' does involve real self-knowledge, a direct acquaintance with one's individual essence or *haecceity*. Attributing something to oneself (for example, some mental state), according to Chisholm, involves being aware of one's own individual essence and seeing that the thing that has that essence (oneself) has a certain property (the state one is attributing to it). "I would conclude," he writes, "that I am able to individuate myself *per se*. I do so in virtue of my awareness of being this particular person. This awareness is a knowledge of propositions implying my individual essence or *haecceity* and is implicit in each of my self-presenting states [such as *my feeling a pain*]. Every such state is necessarily such that, if it obtains, then I am certain of my being this particular person."[40] Thus, Chisholm approaches a substance theory. But he, too, is unaware of the thesis of self-recognition as developed by Butler and Reid, which provides support for his view.

I have not attempted to defend the doctrine of self-recognition in this article, except to suggest that it is not patently untenable. My purpose has been merely to recover the idea, which today remains largely obscured, from the history of philosophy. For even those who have recently worked out positions that are related to this doctrine seem not to have fully understood or appreciated it. I hope that its significance has in some measure been illumined by the comparison with Indian philosophy. At the same time, I hope that the references to Butler and Reid provide more to go by in understanding philosophically this doctrine as it is presented in the Indian texts.

Although, as I have brought out, the original thrust of the doctrine of self-recognition was to establish the permanent, substantial nature of the self or

soul, its relevance to the contemporary debate about *personal identity*, as carried out in abstraction from the issue of substantiality, should be clear. In general, if there is a way of determining that there is a single, continuously existing self, then it ought to translate into a criterion for personal identity. Thus, if the doctrine of self-recognition is true—if we really do re-identify ourselves as the subjects of past actions and experiences—then my remembering *myself* to have done X in the past is a criterion (qua sufficient condition) for knowing that the person who did X was myself; and my identity over time would consist just in the experiences of my various person stages being experiences that *I* underwent. Moreover, if there is such a thing as self-recognition which identifies the self existing now as *strictly identical* with a self that existed before now, then we have reason to believe that a person is not a body or a mind (aggregate of psychological states), since those are things which undergo change over time. A body or mind existing now will not be strictly identical with any body or mind existing before now. Rather, a person must be some other, distinct kind of thing that persists essentially without change over time.

ABBREVIATIONS

MSBh *Mīmāṃsāsūtrabhāṣya* of Śabarasvāmin, ed. E. Frauwallner, in *Materialien zur ältesten Erkenntnislehre der Karmamīmāṃsā*, Österreichische Akademie der Wissenschaften, Philosophisch-Historische Klasse, Sitzungsberichte, vol. 259, Abhandlung 2 (Vienna, 1968).

NM *Nyāyamañjarī* of Jayantabhaṭṭa, ed. K. S. Varadacharya (Mysore: Oriental Research Institute, 1983).

NSBh *Nyāyasūtrabhāṣya* of Vātsyāyana, in *Nyāya Darśana*, ed. Taranatha Nyaya-Tarkatirtha and A. Tarkatirtha (Calcutta: Metropolitan Publishing House, 1936).

NV *Nyāyavārttika* of Uddyotakara, in *Nyāya Darśana*, ed. Taranatha Nyaya-Tarkatirtha and A. Tarkatirtha (Calcutta: Metropolitan Pulishing House, 1936).

NVTṬ *Nyāyavārttikatātparyaṭīkā* of Vācaspatimiśra, in *Nyāya Darśana*, ed. Taranatha Nyaya-Tarkatirtha and A. Tarkatirtha (Calcutta: Metropolitan Publishing House, 1936).

ŚV *Ślokavārttika* of Kumārilabhaṭṭa, ed. Svāmī Dvārikadāsa Śāstrī, Prāchyabhārati Series, no. 10 (Varanasi: Tārā Publications, 1978) (*ātmavāda-adhikaraṇa* unless otherwise noted).

TS *Tattvasaṅgraha* of Śāntarakṣita, ed. Svāmī Dvārikadāsa Śāstrī (Varanasi: Bauddha Bharati, 1961).

NOTES

1. Besides expounding at considerable length the interpretation I am about to consider via Śabara, Uddyotakara also mentions that the *sūtra* can be interpreted to mean that desire, etc., constitute the middle term (*liṅga*) in an inference to the existence of the self insofar as they are qualities which require a self as their substratum (*NV*, pp. 192, l. 10–193, l. 9). (See in this regard Arindam Chakravati, "The Nyāya Proofs for the Existence of the Soul," *Journal of Indian Philosophy* 10 (1982): 211–238, and Klaus Oetke, *"Ich" und das Ich: analytische Untersuchungen zur buddhistisch-brahmanischen Atmankontroverse*, Alt und Neu-Indische Studien, no. 33 (Wies-

baden: Franz Steiner Verlag, 1988), pp. 359–361. The latter work will be the definitive treatment of the Buddhist-Hindu debate about the self for some time to come.) A third interpretation mentioned by Uddyotakara (p. 192, ll. 5–9) is as follows: the inner states of desire, etc., point to a single causal factor (= the self) that is involved in the occurrence of each of them. For they are all connected together (in one consciousness) by virtue of the memory of having been experienced "by me." Similarly, ideas or feelings arising in the minds of various spectators of a dance indicate a single factor in their origination—a particular gesture of the dancer—insofar as they are each accompanied by an awareness of the gesture. (Cf. Śāntarakṣita's discussion of Uddyotakara's position, *TS*, *kārikās* 180 ff.) Vācaspatimiśra points out that this third interpretation renders the *hetu* of the argument an *anvayavyatirekī hetu*, whereas the favored first interpretation, which I consider here, involves a *kevalavyatirekī hetu*. Cf. Oetke, *"Ich" und das Ich*, pp. 355–359.

2. *MSBh*, pp. 52–54, and *ŚV*, *ślokas* 107–110.

3. Neither the *sūtra* nor Vātsyāyana in his *bhāṣya* refers explicitly to memory. Vātsyāyana talks only about the "connecting up" (*pratisandhāna*) of numerous experiences, which he believes indicates a single entity which connects them (*ekam anekadarśinaṃ pratisandhātāram*). (Jha's translation of *pratisandhāna* as 'recollection' seems too specific; cf. Ganganatha Jha, *The Nyāya Sūtras of Gautama, with the Bhāṣya of Vātsyāyana and the Vārttika of Uddyotakara* (Delhi: Motilal Banarsidass, 1984), vol. 1, pp. 217–218. For Kamalaśīla, commenting on *TS* 180, notes that *pratisandhāna* is simply the combining of various cognitions as a result of their having a common cause: *pratisandhānam ucyate yad ekam arthaṃ nimittīkṛtya pratyayānāṃ sambandhanam*. This cause may be simply the experience object. Thus, the cognitions of various spectators of a dance are "connected" (*pratisandhīyante*) through their common experienced object, the dance. See note 1 above; see also Vācaspatimiśra, *Yogasūtrabhāṣyavaiśāradī* 4.15: *abhedaś cārthasya jñānabhede 'pi pramātṛṇāṃ parasparapratisandhānam avasīyate. asti hi raktadviṣṭavimūḍhamadhyasthānām ekasyāṃ yoṣiti pratīyamānāyāṃ pratisandhānaṃ yā tvayā dṛśyate, saiva mayāpīti.*) But Uddyotakara clarifies, at the outset of his *Vārttika* on 1.1.10, that desire, etc., indicate a substantial self insofar as they occur as "having a single object, together with memory": *yasmād evecchādayaḥ smṛtyā sahaikaviṣayā bhavanti tasmād ekakartṛkatvaṃ pratipādayanti* (*NV*, p. 185, ll. 8–9). That is to say, my present desire and a certain past experience are united not just insofar as they concern the same object, but insofar as *I remember* that the thing I desire now is of the sort I experienced to be a cause of pleasure in the past. This formulation is designed to obviate the Buddhist objection that one has to do here merely with a relation of cause and effect between discontinuous, selfless entities: an experience gives rise, as it were, mechanically, to a desire having the same content, as a seed eventually gives rise to a similar seed (via the growth of a fruit-bearing tree). But the Naiyāyika thinks that memory necessarily brings a single, continuous self into the picture: I cannot *remember* something that someone (or something) else has experienced (see *NV* and Vācaspati's *NVTṬ*, pp. 187–189). Śabara and Kumārila, interestingly, try to do justice to an argument from desire that is logically distinct from an argument from memory, but their efforts are unconvincing (see *MSBh*, p. 54, and *ŚV* 103–105).

Note that if one approaches *NS* 1.1.10 from this angle—i.e., as indicating the self as the element that connects several interrelated states over time—the order of the terms in the compound (still a *dvandva*) may be significant. That is, the compound could be referring to a single series of connected experiences which characterize human action: *desire* or *aversion* towards certain things experienced in the past give rise to the *effort* to obtain or avoid them, which is followed by the *experience of pleasure* (*sukha-jñāna*) or the *experience of pain* (*duḥkhajñāna*) from obtaining the desired or unwanted thing (or else: the *pleasure* and *pain* resulting from *effort* are followed in turn by recognition (*jñāna* = *pratyabhijñāna*) of the thing). Vātsyāyana, however, does not consider the compound as referring to one series in this way. He separates off *jñāna* as referring back to something one was previously curious about, not something desired or unwanted.

4. In Buddhist texts we encounter this theory, e.g., in essence, in the *Abhidharmakośa* discussion of the *pudgalavāda* (ed. P. Pradhan (Patna: Jayaswal Research Institute, 1975), pp. 472–473). The alternative Buddhist theory of memory is mentioned by Kumārila (*ŚV* 104–105). Śabara's rejection of the memory argument is actually much cruder, to wit (in essence): "Who says memories cannot take place without a permanent self? For we observe memories to occur. What we do not observe, on the other hand, is a self. How can you say that what we observe is impossible, but what we do not observe must exist?" (*MSBh*, p. 54). This is certainly very thin!

Since Kumārila recognized the Buddhist rejection of the argument from memory to be based on an alternative account of memory in terms of *vāsanās*, it is puzzling why he does not discuss the strenuous efforts of the Naiyāyikas to refute the theory of *vāsanās* in this connection. (Certainly Kumārila was not merely deferring to the Bhāṣyakāra; he has no qualms about correcting him elsewhere.) Thus, Uddyotakara (ad *NS* 1.1.10) criticizes the Buddhist view on the basis of the impossibility of a *vāsya-vāsaka* relationship between momentary entities (see Oetke's critical evaluation of Uddyotakara's argument here, *"Ich" und das Ich*, pp. 347–352). Kumārila himself appeals to this same idea in his *nirālambana-adhikaraṇa* (*ślokas* 178 ff.) and, with respect to the problem of the continuity of *karman* and the fruit of *karman*, even in the *ātmavāda-adhikaraṇa* (44 ff.). But he never applies it to salvaging the Brahmanical theory of memory.

5. *ŚV* 92.

6. *MSBh*, p. 56: *anyedyur dṛṣṭe 'paredyur "aham idam adarśam" iti bhavati pratyayaḥ. pratyagātmani caitad bhavati, na paratra. paro hy asau yo 'nyedyur dṛṣṭavān. tasmāt tadvyatirikto 'nyo 'sti, yatrāyam ahaṃśabdaḥ. . . na vayam "aham" itīmaṃ śabdaṃ prayujyamānam anyasmin arthe hetutvena vyapadiśāmaḥ, kiṃ tarhi śabdād vyatiriktaṃ pratyabhijñāpratyayam. pratīmo hi vayam imam arthaṃ "vayam evānyedyur upalabhāmahe, vayam evādya smarāma" iti. tasmād vayam imam artham avagacchāmo "vayam eva hyo, vayam evādya" iti. ye cāmī hyo 'dya ca, na te vinaṣṭāḥ.*

7. I have tried consistently to interpret this self-recognition as having as its object the subject of the remembered experience. That is, in this recognition, the (past) subject of the remembered experience is identified or 'recognized' as one's present self. This is in harmony with the formulations of Śabara—*anyedyur dṛṣṭe 'paredyur "aham idam adarśam" iti bhavati pratyayaḥ* (passage cited in previous note)—and Kumārila (*ŚV* 115):

yadi syād jñānamātraṃ ca kṣanikaṃ jñātṛ tatra vaḥ |
na bhavet pratyabhijñānaṃ pūrvajñātari samprati ||

However, this is also somewhat problematic in that recognition is usually defined as the identification of a presently existing thing as something previously experienced (*pūrvajñānaviśiṣṭārthagrāhitvaṃ pratyabhijñāyāḥ, NM*, vol. 2, p. 310, l. 9; *atītāvasthāvacchinnavastugrahaṇam, Tarkabhāṣā*, etc., cited in M. Bhimacarya Jhalakikar, *Nyāyakośa* (Poona: Bhandarkar Oriental Research Institute, 1978), p. 543). Thus, *pratyabhijñā* is usually formulated: *sa eva ayaṃ ghaṭaḥ*, "This pot is indeed *that* one [which I saw before]." Apparently the Mīmāṃsakas were not particularly concerned about whether self-recognition in memory fits precisely the definition of recognition. What they stress is that in self-recognition (however it is to be analyzed) we become aware that "I existed then and I exist now." Thus Śabara in the passage above (note 6) and Kumārila, *ŚV* 116: *jñātavān aham evedaṃ pūredānīṅ ca vedmyaham.*

8. See *NV* ad *NS* 3.1.1, pp. 706–707.

9. In this way, too, Mīmāṃsā parallels Nyāya; for the latter generally argues that the use of the word 'I' corresponds to a self-perception. See *NV* 3.1.1, pp. 704–705.

The Nyāya author Jayantabhaṭṭa, however, rejects this view, which he identifies as held both by Mīmāṃsakas (*aupavarṣāḥ*) and some other Naiyāyikas (*sāyuthyāḥ*) (*NM*, vol. 2, pp. 268–278). The true Nyāya position, according to Jayanta, is that the existence of the self is only inferred. His basic objection to the view that 'I' refers to the self is that it implies that the self is both agent and object in an act of knowledge (*naikasyāṃ pratītāv ātmanaḥ kartṛtā karmatā ca syātām*, p. 271). If 'I' refers to anything, he argues, it refers only to the body. (He also considers the possibility that it is a mere word referring to nothing (*sabdamātroccāraṇam*).) Jayanta also rejects the view—perhaps known to him more from the writings of the Kāshmīrī Śaiva philosophers than the Advaita Vedāntins—that the self is self-luminous consciousness (*citiśaktisvabhāvam aparasādhanam*), for he denies that the self is essentially conscious:

sacetanaś citā yogāt, tadyogena vinā jaḍaḥ |
nārthāvabhāsād anyad dhi caitanyaṃ nāma manmahe ||

("[The self] is conscious due to connection with consciousness; without such connection it is insentient. For we do not conceive of any consciousness that is not the manifestation of an object" (p. 275)).

Jayanta characterizes the Nyāya position he opposes as the view that the self would already have to be known by perception in order to be inferred—and then any inference would be unnecessary. That the self thus becomes an object of knowledge for itself should pose no problem:

tatrānumānajñānasya yathātmā yāti karmatām |
tathāhaṃpratyayasyaiṣaḥ pratyakṣasyāpi gacchatu ||
dehādivyatiriktaś ca yathā liṅgena gamyate |
tathāhaṃpratyayenāpi gamyatāṃ tadvilakṣaṇaḥ ||

("Just as the self becomes the object of inferential knowledge, let it be the object of perceptual knowledge in the form of the 'I' concept! As something distinct from the body, etc., is known via an inferential mark, let the self be known as distinct from the body, etc., via the 'I' concept!" (p. 276)). Also, the opponent asks, how could present and past experiences be assigned to one knower—which is the basis of the Nyāya inference to the existence of the self—if one did not have awareness of the self as the subject of thought throughout those experiences? But, then, the self is already given; it need not be inferred:

jñānecchāsukhaduḥkhādi khiledaṃ liṅgam ātmanaḥ |
ekāśrayatayā jñātam anusandhātṛbodhakam ||
tathātvena ca tadjñānam āśrayajñānapūrvakam |
jñāte tatrāphalaṃ liṅgam, ajñāte tu na liṅgatā ||
tasmāt pratyakṣa evātmā varam abhyupagamyatām |
vṛddhāgamānusāreṇa saṃvidālokanena ca || (pp. 277–278)

In answer to this, Jayanta proceeds to show how the self is inferred by the classic Nyāya *kevalavyatireki anumāna*—*without* prior acquaintance with the self. A study of Jayanta's treatment of the self would be of much merit.

10. My thanks to B. K. Matilal for drawing this to my attention.

11. *NSBh*, pp. 710–712.

12. See Vātsyāyana's bizarre discussion of *NS* 3.1.7 concerning the distinctness of the right and left eyes as faculties.

13. *NSBh*, p. 734.

14. *NSBh*, pp. 734–735.

15. The difference between the Nyāya and Mīmāṃsā positions that I have sought to develop here is contained, essentially, in Kumārila's *śloka* 109:

smaraṇapratyabhijñāne bhavetāṃ vāsanāvaśāt |
anyārthaviṣaye, jñātuḥ pratyabhijñā tu durlabhā ||

This asserts that recognition and memory can be explained in terms of *vāsanās*, without appeal to a self, when it is some other object which is remembered or recognized, but not when it is the self that is recognized.

16. Technically, Śāntarakṣita refers back to the formulation of the Mīmāṃsā argument, in *TS* 238–239 (= *ŚV* 136–137), as follows: "The knower (*jñātṛ*) apprehended in a past 'I' awareness (*vyatītāhaṃkṛtigrāhya*) continues in the present, since it is comprehended by the [same] notion 'I', like the present knower (*idānīntanabodhṛvat*)." I.e., we know that the cognizer of a past experience exists now, because it is referred to by the word 'I,' just as the present cognizer. But, Śāntarakṣita argues, this argument is without a valid *dṛṣṭānta*. No present cognizer is refered to by the word 'I', for the word has no basis! Hence the argument fails. See especially *TS* 283–284, together with Kamalaśīla's commentary.

17. This is, of course, a simplification. Some Western philosophers take the expressions 'person' and 'soul' to be synonymous (e.g., Anthony Quinton, "The Soul," *The Journal of Philosophy* 49 (1962): 393–409).

In Indian philosophy an important secondary consideration is the nature of the self: is it individual or universal—i.e., are there multiple selves or only one; is it essentially conscious or not or, indeed, is it consciousness itself?

18. A 'person stage' is defined by John Perry as a set containing all person-events—whether physical or mental or both—which are held to comprise a given person at a given time (see his *Personal Identity* (Berkeley: University of California Press, 1975), p. 9). A 'criterion' in the

discussion of personal identity is either the ground by which we judge two person stages to be one person or else the feature that objectively constitutes their being such. Perry likens a person stage to the inning of a baseball game. The question, what criterion or criteria establish two person stages to be of one person, is analogous to the question of what criteria establish two baseball game innings to be innings of the same baseball game.

19. H. P. Grice, "Personal Identity," *Mind* 50 (1941), reprinted in Perry, *Personal Identity*, and Anthony Quinton, "The Soul," also reprinted in Perry.

20. John Locke, *An Essay Concerning Human Understanding*, ed. A. C. Fraser (Oxford: Clarendon Press, 1894), vol. 1, pp. 466–467; see the rest of sec. 26.

21. Ibid., p. 465.

22. This ties in with Locke's definition of a person as "a thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places; which it does only by that consciousness which is inseparable from thinking, and, as it seems to me, essential to it: it being impossible for any one to perceive without perceiving that he does perceive" (ibid., pp. 448–449).

23. Ibid., p. 451.

24. Ibid., pp. 450–451.

25. Joseph Butler, *The Analogy of Religion* (New York: Carter and Brothers, 1936), pp. 299–300. Howard M. Ducharme has recently demonstrated that the position I discuss here in relation to Butler and Reid was first set out by Samuel Clarke: "Personal Identity in Samuel Clarke," *Journal of the History of Philosophy* 24 (1986): 359–383.

26. Thomas Reid, *Essays on the Intellectual Powers of Man* (Cambridge, Massachusetts: MIT Press, 1969), p. 341. Cf. also p. 326: "The remembrance of a past event is necessarily accompanied with the conviction of our own existence at the time the event happened. I cannot remember a thing that happened a year ago, without a conviction as strong as memory can give, that I, the same identical person who now remember that event, did then exist."

27. Butler, *Analogy of Religion*, pp. 302–303.

28. Cf. Reid, *Essays*, p. 359: "It is very true, that my remembrance that I did such a thing is the evidence I have that I am the identical person who did it. And this, I am apt to think, Mr. Locke meant: but to say that my remembrance that I did such a thing, or my consciousness, makes me the person who did it, is, in my apprehension, an absurdity too gross to be entertained by any man who attends to the meaning of it: for it is to attribute to memory or consciousness, a strange magical power of producing its object, though that object must have existed before the memory or consciousness which produced it."

29. Butler, *Analogy of Religion*, pp. 302–303. In view of this and other passages it is puzzling that Butler and Reid have been considered critics of the memory criterion of personal identity *on the grounds that it is epistemically circular*—i.e., that memory is evidence for an event only if it is veracious, whereas our confidence that memory in a particular case is veracious must be based on our already knowing by some other means that the event actually occurred. Terence Penelhum, "Personal Identity," in *Encyclopedia of Philosophy* (New York: Macmillan, 1967), vol. 6, p. 98, mentions Butler as the originator of this argument, referring to the oft-cited passage (Butler, *Analogy of Religion*, p. 298): ". . . consciousness of personal identity presupposes, and therefore cannot constitute, personal identity, any more than knowledge, can constitute truth, which it presupposes." Here, however, Butler is merely making the same point as Reid in the passage cited in the previous note, that memory does not *make* someone the same person. For both Reid and Butler, memory—i.e., self-recognition—provides indisputable evidence of personal identity (one's identity with a previous person stage), but it cannot constitute personal identity. The latter must already exist as a fact for memory to enable us to ascertain it.

30. The Buddhist response to this is, first, to imply that a memory impression (*vāsanā*) can indeed be a cause of error: Kumārila himself, Śāntarakṣita charges, holds that the various notions about God—that he is creator of everything, omniscient, etc.—conceived by devotees of God are merely errors rooted in *vāsanās* (?; *TS* 282; I am unable to locate anywhere in the *ŚV* where Kumārila makes such a claim). Of more consequence is a general critique of *pratyabhijñā* as a *pramāṇa*, presented in *TS* 444–460, in the context of the discussion of the theory of the momentary existence of entities. The Mīmāṃsaka holds that the recognition of objects from the past

directly refutes the *kṣaṇabhaṅgavāda*. Śāntarakṣita defends it by charging that *pratyabhijñā*, recognition, cannot be considered a *pramāṇa* qua perception, for it is conceptual (*TS* 446) (whereas according to the Buddhist definition of perception it must be void of concept, *kalpanāpoḍha*, the reality of things being inexpressible), and it is (sometimes) erroneous: e.g., people believe they 'recognize' hair or grass that is actually new hair or grass that has grown back after being cut (*TS* 447, 450; see *Pramāṇavārttika, pratyakṣapariccheda* 503–505 (Varanasi: Bauddha Bharati, 1968)). Moreover, *pratyabhijñā* simply re-cognizes an object that was previously known by another means of knowledge, as does memory. Thus, it does not function as the most effective factor in the cognition of the thing, and so is not a *pramāṇa* (*TS* 451; Kamalaśīla clarifies that this argument appeals to the notion of a *pramāṇa* as the *sādhakatama-kāraka* in the production of a cognition; cf. *Pramāṇavārttika, pratyakṣa* 506). The reader is invited to peruse for him- or herself Kumārila's response, *TS* 452–456 (again, these *kārikās* are not identifiably from the *ŚV*), and Śāntarakṣita's reply, *TS* 457–460. In essence, Kumārila argues that *pratyabhijñā* is a valid means of knowledge, a *pramāṇa*, qua perception. The debate about *pratyabhijñā* is also carried on in the *Nyāyamañjarī*. Contra Mīmāṃsā, Jayanta does not believe that recognition establishes the permanence of letters and words (*NM*, vol. 1, pp. 554–557), but, now with the Mīmāṃsaka but against the Buddhist, he believes that other kinds of objects are indeed recognized (*NM*, vol. 2, pp. 307–314, 328–335).

31. Another interpretation of this idea (that of Umbeka, an early commentator on the *ŚV*) is that the validity of a cognition is produced just by the factors—object, sense faculty, etc.—that produce the cognition: "the causes of the cognition bring about validity" (*tasya [prāmāṇyasya] jñānahetava evotpādakāḥ, Ślokavārttikavyākhyātātparyaṭīkā*, ed. S. K. Ramanatha Sastri (Madras: University of Madras, 1971), p. 53, ll. 25–26). If, explains Pārthasāthimiśra in his discussion of this view in his *Nyāyaratnamālā* (ed. Rамanatha Sastri, Gaekwad Oriental Series, no. 75 (Baroda: 1937), p. 48), the faculty of cognition, etc., could produce a valid cognition only if endowed with some excellence (*guṇa*) (the Nyāya view), then there would be no validity at all to cognitions arising from defective faculties, etc. But even the cognition of a yellow conch produced by a jaundiced eye has a correct aspect—the 'conch' aspect. Thus, every cognition, insofar as it merely arises from "its own cause" (*svīyāt kāraṇāt*), i.e., from the factors that produce the cognition itself, has some truth to it. And a cognition that arises from factors that are free of defects will be completely true. (Pārthasārathi, however, goes on to dispute that this is a correct interpretation of Kumārila's statements on *svataḥ prāmāṇya*).

32. A definitive study of the *svataḥ prāmāṇya* doctrine has yet to be written, but much useful information about it is contained in L. Schmithausen, *Maṇḍanamiśra's Vibhramaviveka*, Österreichische Akademie der Wissenschaften, Philosophisch-historische Klasse, Sitzungsberichte, vol. 247, Abhandlung 1 (Vienna, 1965). See, e.g., pp. 189–201. This doctrine is of great relevance to contemporary epistemology. See, e.g., the attempt by Nicholas Wolterstorff to revive Reid's epistemology in the service of the anti-evidentialist defence of theistic belief: "Can Belief in God Be Rational?" in *Faith and Rationality*, ed. A. Plantinga and N. Wolterstorff (Notre Dame, Indiana: University of Notre Dame Press, 1983).

33. "Personal Identity and Memory," in Perry, *Personal Identity*, p. 126.

34. Sydney Shoemaker, *Self-Knowledge and Self-Identity* (Ithaca, New York: Cornell University Press, 1963), pp. 162–164.

35. On self-acquaintance see *MSBh*, pp. 56–58; *ŚV* 142. The Mīmāṃsā position is that the self is known or revealed by itself (*svasaṃvedya; ātmanā prakāśya*). This poses a certain problem for the Mīmāṃsaka, however. He wishes to deny, against the Buddhist idealist, that cognitions (*jñāna*) can experience themselves; for one thing cannot function both as subject and object (*karman*) in the same act. Thus, one thing cannot be both perceiver (*grāhaka*) and perceived (*grāhya*) in an act of cognition. But then how can the *self* know itself without violating this principle? Kumārila's answer, offered at *ŚV, śūnyavāda* 68–71, is obscure. I decipher it as follows: The cognition (*jñāna*) by which the self knows itself is a property (*dharma*) of the self, while the self is known as "substance, etc." (*dravyādi*). Thus, we do not really have a case of the same thing knowing itself. Rather, we have one aspect of a thing knowing another aspect of the thing. But even if one assumes that in the case of the awareness 'I' one has to do with a cognition knowing itself (*jñānasyaiva. . . saṃvittiḥ. . . ātmakartṛkā*), that would be permissible here, since

the consciousness of 'I' arises with respect to an agent factor (*kartari bhavanti*); i.e., it has as its content the knower (*jñātṛ*), which conceivably could be the cognition (s. 70, according to Pārthasārathi's commentary). But that is not permissible for any other kind of cognition—a cognition of blue, e.g., since one is only aware of the object 'blue'. In other cognitions one is not aware of the knower.

36. Shoemaker, *Self-Knowledge and Self-Identity*, pp. 239–243. Note, however, that Shoemaker develops a rather different account of the non-criterial nature of self-identification in his later article, "Persons and Their Pasts," *American Philosophical Quarterly* 7 (1970): 269–285.

37. Geoffrey Madell, *The Identity of the Self* (Edinburgh: The University Press, 1981); see especially pp. 122–127.

38. Ibid, chap. 2, pp. 23–48.

39. Ibid, pp. 124, 134–138.

40. Roderick Chisholm, *Person and Object* (La Salle, Illinois: Open Court Publishing Co., 1976), p. 37.

This page intentionally left blank

I Touch What I Saw

ARINDAM CHAKRABARTI
University of Washington.

> "Like a viewer standing behind many windows there must be someone the same who sees both"
>
> —Praśastapāda[1]

> "...if only a person can have illusions it cannot be an illusion that some person exists."
>
> —Geach[2]

> "As long as the idealist stands unrefuted, not even the external world exists, so where is the place for the self?"
>
> —Udayana[3]

> "What would this whole play be without an onlooker?"
>
> —Frege[4]

0. Introduction: Realism-links

Realisms about the self and about the external world entail each other. And both realisms derive support from the plain fact of perceptual re-identification of objects across times and senses. That is going to be the major contention of this paper.

The Berkeleyan view that there is an abiding ego undergoing subjective states but no material objects for these states to focus on will be shown to be inconsistent. Equally incoherent, I shall argue, is the Parfitian view that there are more or less persistent material objects, e.g., buildings in Venice and hemispheres of brains, and also experiences and thoughts about them but no stable owners of those states of consciousness. A reductive anti-realism about selves will logically commit us to a similar diffusion of material bodies into their secondary qualities, and of those, in turn, to subjective

---

1 *Padārthadharmasaṃgraha*, Chapter on the self as a substance.

2 *Truth, Love and Immortality*, p. 38 (Berkeley and Los Angeles: University of California Press, 1979).

3 *Atmatattvaviveka* (see reference in note 6) p. 428, my translation.

4 *The Thought: A Logical Enquiry*, p. 32 (see note 7).

experiences and expectations. If we are afraid of such idealisms we should better not flirt with such "self"-lessness.

That seems to leave us with only two alternatives: Either embrace dualism, i.e., a full-fledged realism about substantial selves and material bodies or adopt a kind of radical empiricism reducing both egos and objects to nebulous aggregates of twilight qualities. Nyāya and Descartes took the first option. Buddhists and James went for the second. Perhaps there are other more drastic revisionary alternatives. The views I consider all accept the reality of thoughts or experiences—the common middle ground uncontroversial across the board for a dualist, idealist, no-soul-ist, radical empiricist or even Lichtenberg. Of course one could deny even that and just consider matter to be real. In *this* paper I cannot argue in detail against all these eliminative reductive views. But I don't want to end up in a Jaina vein with a disjunctive metaphysics.[5] It will be clear that the common-sense realistic package is more to my taste. Against the picture of us all as ownerless object-less loosely continuous rivers of consciousness, I should like to argue that it fails to account for the kind of re-identificatory experience to which the title of this paper gestures. By eventually fudging the distinction between one person and another, such reductionism makes the private-public, subjective-objective divides fuzzy. That results in making linguistic communication unintelligible. Plain facts evaporate, licenses to make one's own world are issued while the supposed worldmakers themselves keep dividing, merging, re-emerging and partially surviving. To the extent that we conceive of ourselves that way, *cares* would lessen, no doubt, but so would *caring*. Maybe we shall be liberated from egoism to some extent. But whose liberation will it be anyway?

1. The Method: Good old Nyāya Wine in a Fregean Bottle

In building up my case I shall consider realistic arguments from both classical Indian and modern Western sources. Udayana—the 11th century precursor of the Indian 'New Logic' (Navya-Nyāya)—saw the connection among realisms about physical objects, selves and universals, and defended all three against the Buddhist reductionists who deemed all three to be grammatical fictions.[6] Frege, the father of contemporary analytic

---

5 The basic Jaina methodological framework is one of non-sceptical and non-agnostic alternation. Among conflicting theories of reality t_1, t_2, t_3 etc., Jaina philosophy recommends a reconciliation roughly in the following form, "Either t_1 (from point-of-view v_1) or t_2 (from point-of-view v_2) or t_3 (from point-of-view v_3)...etc." There is no picture of any exclusive ultimate reality inaccessible to and beyond all these alternations. Hence my phrase "disjunctive metaphysics." The disjunction is not due to epistemic limitations but lies at the very heart of reality which is non-one-faced (*Anekānta*).

6 See '*Ātmatattvavivek*' by Udayana-ācārya, edited by Laksmana Śastrī Dravīda, The Asiatic Society Calcutta, originally published in 1907–39, reprinted 1986. A not too accurate

philosophy, felt the need to establish the self as a bearer of ideas in order to secure the objectivity of 'thoughts' of which facts are a species.[7]

There are important differences between the Nyāya and the Fregean ontologies. Their concerns too were different. But both seem to recognize the necessity and connectedness of subjective and objective unifiers of experience, the inter-dependent double need of mental states to be *someone's* and *of something*. By attempting, in my own way, a defense of a rather dated dualism of real selves and real objects, I hope incidentally to re-explode the myth that all Eastern Philosophy is idealistic, monistic and mystical. Post-Parfitian Western philosophy will also, I hope, have more to learn from the East than mere mindlessness.

Although I shall chiefly deal with the basic insights and arguments, ignoring history and hermeneutic, I hope the exercise will illustrate the point that the reality of a philosophical problem is proved by its re-identification across centuries and cultures. The mapping of perceptual states onto the person and the external environment raises problems that I *saw* in Gautama's 2nd century aphorism[8] and can now *touch* again in Colin McGinn's 1989 work on mental contents (which, however, upholds a different view of persons and objects).[9]

2. The Ego and the It

I am going to argue for the truth of a bi-conditional: Selves over and above experiences are real *if and only if* objects over and above experiences are real. In this section I shall try to prove the '*if*'-part of this claim: If we are realists about the external world then we have to introduce substantial egos into our ontology.

Now, to be a realist about material objects, let's say an apple, is at least to uphold the following two strict distinctions and the two strict identities:

(A) The apple is distinct from just an experience or idea of an apple.

---

summary of the work is to be found in Karl Potter (ed.), Encyclopedia of Indian Philosophies (Princeton), Vol. II (Nyāya Vaiśesika), pp. 526–56. I have drawn heavily from especially chapters 3 and 4 of this vast work entitled, "The dispute about denying the distinction between property and property-possessor" and "The dispute about non-apprehension of the self"—respectively.

7 See "*The Thought: A Logical Enquiry*" by G. Frege translated by A and M Quinton, in Strawson (ed.), *Philosophical Logic* in the Oxford Readings in Philosophy series. 1967.

8 "Darśana-sparśanābhyām ekārtha grahanāt"—Nyāya-Aphorisms, No: 3.1.1. The exact context is the proof that any individual sense-faculty (or their conglomeration) cannot serve as the self. Literally translated, it means, "Because the same object is grasped by both sight and touch."

9 McGinn, Colin, *Mental Content*, Oxford 1989.

(B) The apple is distinct from just its red-color, its coldish touch, smooth texture and sweet taste, and also from just a series or set of these qualities.

(C) The apple that I now touch (or bite or smell) is the same as the apple that I saw a while ago.

(D) The apple that I see is the same as the apple that you see.

We can tease out the existence of a self straightaway from *A*—the first distinction—because it implies the reality of both the items distinguished, viz., the apple and some experience of it. And experience cannot exist without an experiencer. The asymmetric individuation-dependence of experiences on relatively permanent subjects or owners of experience can be brought out at four different levels, viz., linguistic, epistemological, metaphysical and conceptual.

> Linguistically, talk of perceptions presupposes prior reference to perceivers.
>
> Epistemologically, no sensations are knowable directly or indirectly without some sort of knowledge of the person who senses.
>
> Metaphysically, it seems ontically essential to a sensory state that it be unsharably borne by a conscious individual (like a dent needs a dentable surface).
>
> Conceptually, we cannot conceive of a token experience without having the concept of a bearer of that experience.

Notice that each of this fourfold identifiability-dependence is only one-way. For instance, take the epistemological dependence of mental states on the self. It might seem that conversely even a self cannot be known in isolation from a mental state. But a self's individuation-dependence is not on any *particular* token-experience but at most on some experience or other; whereas a particular token experience, if it has to be known, has to be known along with that particular subject to which alone it can be ascribed and not with some subject or other.

Some Buddhists insisted that the so-called self can be *reduced*[10] to the individual mental qualities with which it is constantly co-cognised. But even if symmetric constant co-cognisability proved identity, one-sided

[10] For the most reliable and accessible account of this Buddhist view, see Steven Colins "*Selfless Persons*" (Cambridge University Press, 1985).

knowability-dependence surely does not. To use Udayana's counter-example, a dark color cannot but be perceived along with some bright color (e.g., of sun rays or of light in general) yet they are quite distinct, because the bright color *can* be seen without (be seen without) the dark.

Could it be that although a single passing thought is not the thinker—the entire stream of thoughts—a seamless series of experiences assumes that office or title of 'self'? In order to be compatible with a realistic view of objects as factorized above, such a series will of course have to be more well-integrated than just a bundle of visual, tactile, olfactory and similar sensations. The temperature cannot be an object of visual sensation and the color can never be an object of tactile sensation (except in special situations like synaesthesia when it is assisted by memory of the very sort of trans-sensuous re-encounter which we are trying to explain). How then can the identity *C*—between what is seen and what is touched—be maintained? It will not do to give the ruling that *all* such claims of re-encountering the same object across times and senses are uniformly illusory. Who commits the mistake? Can just a visual experience so much as imagine that *it* touched a cold surface? We cannot even credit the *pair* of sensations with the unitary conception of what it is for one object to have both color and temperature. A Jamesean picture of one thought dying away and the replacing thought finding it 'warm' and greeting[11] it as "mine" suffers from a number of conceptual flaws. Would such unowned thoughts which 'appropriated' the predecessors only in a rhetorical sense of 'possessing' ever be able to form the notion of mind-independent objects bearing properties accessible to different senses? We shall come back to this question in a moment. Let us pause here to look harder into this notion of *owning an experience.* Vasubandhu while explicating the Buddha's doctrine of *no-soul* faces the same problem;

> How is this locution "Chaitra remembers" possible? If there is no self whose is this memory? Why is the possessive case used here ("whose")? To express the relation of ownership, just as in "This is Chaitra's cow." How is he the owner of the cow? Because the employment of the cow in carrying loads or milking etc. is under his control.
>
> Towards what end is a memory supposed to be employed so that we are looking for its owner? Of course, towards the object remembered.... But how exactly is the recollection to be 'employed'? Is it by production or by movement? Since memory is incapable of movement it must be through production. So we conclude that the producer or the cause is the so-called owner and the effect produced is the owned possession.
>
> ...Hence whatever is the cause of memory is the owner of the memory.

---

11 James, W. '*Principles of Psychology*' Vol. 1, Chapter X, p. 32 (the works of William James, Harvard University Press, 1981).

Add to this the Buddhist[12] notion of a cause as a perishing predecessor, and you get the thesis that any immediate *predecessor* of a mental state is its owner. This version of the Stream-Conception is, in a way, directly opposed to the Jamesean version that the *successor* thought calls its dying ancestor "mine" and "appropriates" it.[13] However, Both doctrines are equally counter-intuitive. Strangely enough, upon one prevalent version of Buddhist epistemology, one aspect of the preceding mental state is also the object (the *ālambana*), the causal and intentional support of the succeeding mental state. So, here we seem to have a peculiar collapsing of the owner and the object of the cognitive state. That may not daunt the Buddhist idealist who professes the doctrine of reflexive self-awareness of individual perceptual states. But it is surely incompatible with realism about the object as spelt out by our initial fourfold list of distinctions and identities. Three (*A*, *C* and *D*) of those claims entail the existence of mental states which need to be ascribed to something. *C* and *D* entail re-identification across times and senses which requires that the separate re-identifier endures in those times and runs across those sensations, because the sensations themselves do not, as it were, speak each other's language in order to even *ask* the question "Have you grasped the same item as I have done?" Even a single experience has the ascriptive demand which is strengthened by the integrative and retentive demands of synchronic and diachronic syntheses of such experiences. These are all demands for the same subject or substance. Suppose we ignore these demands and diagnose this adjectival understanding of mental events as a deep-rooted linguistic prejudice. I hope to show that the revised notion of the self as a heap or stream of mental events will then be inconsistent with distinction *B*, between properties of physical objects and the objects exemplifying those properties. Nothing short of a substantial self is thus implied by that crucial constituent of realism about physical objects.

Suppose, at any single cross-section of time it is a *heap* of perceptual, conceptual, dispositional and hedonic properties dependent upon each other and on correlated bodily and cerebral events; and over a period of time it is just a *stream* of such aggregates maintaining some sort of continuity and insulation. Can we go ahead with such a set and series (*rāśi-dhārā*)[14] notion of the self while retaining the notion of sweet red apples individuated in their environment?

---

12 *Abhidharmakoṣa* by Vasubandhu, Chapter VIII, p. 1217 of the Bauddhabharati Edition. For Theodore Stcherbatsky's translation of this passage see pp. 52–53 of "The Soul Theory of the Buddhists," Bharatya Vidya Prakasan, Delhi, 1976.

13 *Principles of Psychology*, Vol. 1, p. 322, op. cit.

14 *Abhidharmakosa*, p. 1205. "Therefore, the person is only nominally real, like a 'heap' or a 'stream', etc."

If I am just such a club of say visual and gustatory sensations what would the apple be for me? Since I shall not have the framework of *predication* or *exemplification* at my disposal (my self-knowledge being essentially additive), it will be at most a color plus taste or a color followed by a taste. But distinction *B* tells us that the apple is more than just a sum of its qualities. A bundle-self will look upon objects as similar bundles of properties severally accessible to its own members. Indeed, it is hardly intelligible how a 'set-series' sort of self can even put together discrete sense-data as *bundles*. Temporal proximity will hardly do as a deciding factor. While *looking* at the ocean, one may *smell* a friend's cigarette-smoke and *touch* a beach towel. The resulting bundle of contemporaneous visual, olfactory and cutaneous sense-data will yield no recognizably single physical object. Even if by some fluke the last bit of a time-slice of a subject-series succeeds to tie up the appropriate bundle, can it ascribe all the sensible qualities (or even one of them) as *properties* belonging to an external *substance*? Under the assumption that the subject of experience *really* is a series of mental states, one could imagine two alternative pictures. The *first* picture goes like this: The stream-self originally feels its own states in feature-placing forms like "Here comes a feel," "Now, there is a pinch of jealousy," "Here comes a wish," rather than "I touch," "I am jealous," or "I wish." But the same subject feels external sensations in predicative forms like "The apple is red," "The water is cold," "A stone rolled off." The *second* picture removes this incredible asymmetry. Both inner and outer experiences come through predicative judgements; yet, it insists, the subject-predicate model is *mis*-applied to the inner and *correctly* applied to the outer sensations. It should be obvious that both the pictures are equally bizarre. It is extremely arbitrary to hold that while the ascriptive model is *never* or *never correctly* applied to the inner realm, the additive model is *never* or *never correctly* applied to the outer. Experiences of the two realms come to us at the same time, often indistinguishably. Interpreting two aspects of them in two radically different styles leads to either false psychology or false metaphysics. Any such negation of a substantial person is thus incompatible with realism regarding objects. If there is an "*It*" beyond experiences and qualities, there must also be an "*I*" behind them.[15]

---

15 One might ask, "Must this ego be a *substance*?" In Kant the transcendental subject never becomes an object of knowledge, hence is never categorisable as a substance. But I see no reason to feel constrained by such Kantian strictures. All that it takes to be a substance are

(a) ability to bear several qualities
(b) persistence across several moments of time
(c) remaining the same while the qualities change.

Finally, it can be shown that identity *D* of objects of two distinct owners of experience requires the existence of more than one selves. Why can't "you" and "I"—in that part of the realist's claim—just mean rivers or clubs of unowned mental states? Now, rivers can branch off and clubs can exchange members or live intermittently. Such also are Parfitian persons. But two such persons cannot talk of seeing the same physical objects from unsharably distinct perspectives. Since such persons might have merged with one another and could have been teletransported or replicated at any point in their lives *—sometimes sharing the same experiences and memories—* checking against each other's memory for any two of them will be as powerless against mistake as the same person's remembering something twice (recall Wittgenstein's analogy of buying two copies of the morning paper).[16] So those many people whose ability to coperceive the same object is part of the realist notion of the external world, must be substantial selves rather than Parfitian persons. A single Parfitian person does not remain strictly the same over time. Hence two of them would not be strictly distinct. The publicness criterion of objectivity of physical items requires being the common target of many *strictly* distinct centers of experience.

3. Backtracking: Someone In Here Entails Something Out There

I have tried to argue that a realist conception of material objects over and above their properties and independent of our experience of them compels us to believe in the existence of substantial selves, in more ways than one. The other side of the bi-conditional remains to be established. Before we pick that up I want to raise and answer one possible objection. In this section I am going to argue that if we are realists about selves we have to be realists about the external world. But can one even coherently talk about ***realism*** about the self? Can the mind be claimed to be mind-independent? Isn't that why no primary/secondary distinction has ever been drawn about the qualities of the mind? An object can be conceived to bear a quality, especially a primary quality without being recognized by anyone as having that quality. But is it possible for a self to have a state of consciousness without the fact being known to that self? Aren't all perceptions automatically apperceptive?

There are two grounds for this objection. The first is a confusion between "mind" and "self"—which always bothers translators of Indian thought into English. The mind cannot be independent of the mind but the self can. This sounds like a howler to the Western philosopher because not only is the

---

My arguments from the identities and distinctions on the objective side of a series of perceptual experiences prove all these features for the required subject. Hence I do not overstate my case.

16 Remark 265, p. 94c of *Philosophical Investigations*.

distinction between inner sense—the faculty of introspection (which Indians call "Manas" and popularly translate as mind) and the Ego or Self (which they call "ātman") not made in the West (except, in a way, by Kant), but the notion of an objective self—a self which exists even during the gaps when it is not aware of itself—is nearly absent there. The second ground for the objection is the widespread belief that mental states are self-intimating, necessarily recognized by the subject who bears them. The famous "knowledge implies knowledge of knowledge" thesis is a corollary of this. But the Nyāya-epistemologists rejected this doctrine of necessary self-awareness of all states of mind.[17] The self's normal stance is world-ward. Many of its own states may go unregistered if it does not pause to look within and perceive itself as undergoing claims of identity and distinction:

(A′) The subject which ascribes to itself a set of cognitive, conative and affective states at t_1 and the one which ascribes to itself another completely different set of such states to itself at t_2 are strictly the same substance even if the mental states have radically different contents *if* the latter can remember the contents of the former set of mental states from an 'inside point of view' as having been its own. (Notice that such memory is a sufficient but not necessary condition of such identity across times.)

(B′) If a tactile sensation at a subsequent time t_2 immediately arouses the perceptual judgment about a visual sensation at an antecedent time t_1 remembered from inside—as both sensations of the same object—the subjects of these sensations must be the same.

(C′) This subject is *distinct* from any particular visual or tactile experience which inheres in it (it is surely not just the organic feeling of breathing,[18] as James once suggested).

(D′) It is also distinct from the entire set of sense-experiences, volitions, memories and organic feelings at one time or the series of such states across a stretch of time.

---

17 I cannot here go into a full defense of the realists' rejection of the doctrine of reflexivity of all cognitive States. One obvious argument is the infinite regress which will set in if every cognition of x has to be a cognition of cognition of x. The other peculiarly Nyāya argument is that at the same point-instant the self can be in touch with only one (simple or compound) object. Finally, memory does not register every cognitive episode that happens to us, hence, not all of them are even internally apperceived. But this is a subtle and demanding topic of controversy.

18 See James, *Principles of Psychology*, Chapter X, Vol. I.

Notice that each of these identities and distinctions stands on the notion of disparate experiences that the self ascribes to itself and the contents of which it can compare, contrast, link up and look back upon. Now, every experience, even an organic sensation of pain or pleasure, reaches outside itself to get at some physical item or other: a visual sensation individuates itself as of a color, a pain as inside the stomach, a thought as about a picture, and so on. Most of these experiences, especially those which we can self-consciously ascribe to ourselves, have a predicative content where *something* is felt to be of *some sort*, my pen is seen to be blue and shiny or the ice-cream to be melting and sweet. If all we had in our world were myself and my mental properties then either all experienced properties will have to be ascribed to myself, resulting in unusual feelings like I am blue and shiny or I can taste myself as melting and sweet; or, alternatively all our perceptual contents will be just aggregates of qualities conjoined in a heap or series. My experience of a heavy shell will then be reduced to an experience of which one *part* was the sensation of weight and the other part or the whole itself an idea of the shell. But as Frege pointed out, what the experience *claimed* to grasp was not a shell which has weight as its *part*, but one which weight as its *property*. A self which never really comes across an external world can hardly be imagined to have even the idea of objects unlike itself bearing properties unlike its own properties. Indeed, such a self cannot even distinguish between different mental states in terms of the objects they grasp, being deprived of the notion of such objects.

Can such a self still manage to persist across its objectless mental states (which somehow manage to differ phenomenologically!)? Here I borrow an argument from Strawson to show that we cannot think of such a self, or at least explain how it can ascribe differing experiences to itself. To make sense of our trans-temporal identity-claims about the self, we have to appeal to the idea of an objective time-order. Such an order of succession of events will have to be understood as a system of irreversible temporal relations—independent of the self's subjective feelings. Since all that our self can feel are its own experiences, some of those experiences will have to be construed by the self as experiences *of* things outside the experiences. But there is only one way in which perceived *things* can provide us with the background of an objective time-order, namely, "by *lasting* and being re-encounterable in temporally different perceptual experiences."

Udayana anticipates the basic insight of this very argument when he argues just from what I have called identity *B´* to the object of perception being real, mind-independent and irreducible to its qualities or a set of them. He enriches the argument further by bringing in the difference of sensory qualities along with difference of times, going into further details about the relation of inherence which is testified to by our perceptual awareness of the

color being *in* the object. In resisting the idealist pressure to reduce the perceived single object with many features into a mere collection of those features, he also anticipates Frege's point that if there are no objects outside my ideas then the logical distinction between *properties* and *parts* will vanish. Thus it can be shown that a reductive anti-realism about the external world will lead to unavailability of the notion of a self which can re-identify itself across different successive experiential states.[19] Transpose this result and you get the conclusion of this section: *realism about the self implies realism about objects outside the mind.* A persistent self needs to discriminate between sensory states across which it has to ensure its own identity. Sensory states, to be discriminated, need to refer their contents to the external world. Nyāya, like contemporary externalists, holds that individuation of discreet co-personal mental states is to be done directly in terms of mind-independent objects and their equally mind-independent relations. There is no telling the difference between experience as of a blue pot and experience as of a red shirt unless there are blue and red pots and shirts at some time somewhere in the world (if not then and there). The conception of a subjective cognitive state becomes available to us only in the wake of error. Contentful experience spontaneously takes the form "*This pen is red*" and does *not* take the form "I am having a visual experience as of a red pen" or even "Here is an experience as of a red pen."

It is only in the face of suspicion or detection of error that we curtail the claim and become aware of our beliefs. There is a certain cognitive *bashfulness* about confessions of a subjective state of belief which is appropriate only to a believer who fears a mistake.

But the concept of error is patently parasitic on the concept of getting it right, and *rightness* can only be understood in terms of an order of things as they are independently of our beliefs, *that is*, of things which can be common persistent objects of *many* sensations of *many* people on *many* occasions. Since thus, there is no notion of belief without a notion of *things* about which belief can be wrong and right, there is no notion of a believer also without some notion of such outer topics of beliefs.

4. Conclusion: Externalism without Qualms

Our actual—if you like, *naive*—conception of a bit of experience with a structured content involves at least two types of items *outside* the

[19] I had imagined Berkeley to be the exemplar of the position I am attacking by this argument: that external objects are reducible to series of sense-qualities but the self remains substantial. But even Berkeley, I recently noticed, in his consistent moments, goes to the Humean length of asserting: "Mind is a congeries of Perceptions," and denies any real distinction between "that thing which perceives" and "Perceptions" themselves. See remarks 580 and 581 in his philosophical Commentaries, Notebook A.

experience, at its two opposite extremities, as it were. On the one hand, it requires a single experiencer capable of retaining its felt identity over a series of successive experiences. On the other hand, it depends upon an object with some features or a constellation of objects and features in terms of which its content is to be cashed. This dual individuation-dependence makes recognitive perceptual judgements at once evidences of the same I who touches and sees as well as of the same object which is touched and seen. It also conceptually requires that we distinguish the I from the touching and seeing as well as the object from the texture and color.

In case all of the above sounds like just a rehash of the Kantian theme of the objective and subjective unities implied by the synthesis of experience, I wish to emphasize the vital difference. It is crucial for Kant that the *I* which thinks and unifies experience by applying concepts remains only a presupposition of knowledge and never becomes an object of knowledge. This unity of apperception is surely not the empirical self whose knowledge requires representation of something stable in space. This Kantian doctrine of the unknowability of the real subject was fondly picked up by Indian Neo-Vedantists early in this century. K. C. Bhattacharya remarked that if the self is what the word "I" expresses then "it is not meant or at best meant as unmeant and is accordingly above metaphysical dispute."[20] This opinion is by no means shared by the realist Nyāya philosophers whose chief 'metaphysical dispute' with the Buddhists was about the self. Neither is this attitude shared by Frege who, though aware of the incommunicable residue of the 'sense' of our individual uses of the pronoun "I," surely does not think that self or bearer of ideas is not *known* or not *meant* or *referred* to at all by words. Indeed Nyāya believes in the direct perceivability of the real self as much as it believes—once again *unlike* Kant—in the direct knowability of the real world of objects as they are in themselves.

An alleged Humean failure to perceive any self apart from fleeting experiences is symptomatic of the anti-realist. In response to a similar argument from non-perception to nonexistence coming from the Buddhist, Udayana asks: "Is it only you alone who fails to find the self or everyone? You could never be sure about what all others do or do not perceive. As to your own non-perception that is no invariable sign of the self's nonbeing."[21] What then is the positive evidence for the existence of the self? Udayana's answer is unambiguous: "Direct perception, to be sure." For every conscious creature the judgement "I am" is self-evident. This judgement is prelinguistic, so it could not be purely fictional or object-less: there is no question of our being merely duped by the pronoun. Since it is never

[20] *The Subject as Freedom*, in *Search for the Absolute in Neo-Ved ānta* (The University Press of Hawaii, 1976) p. 93.

[21] *Ātmatattvaviveka* of Udayana, p. 738.

challenged by any subsequent judgement it is not doubtful also. It does not await our search for a premise or reason. Hence it is not inferential.[22] It could not always have been just a memory unless at some time it was presented as a datum. So it must be an indubitable perception. Even if we sometimes are mistaken about our own present mental states, a self-delusive judgement like: "I am in pain" or "I *see* a snake"— in order to be false, presupposes the existence of and my direct acquaintance with the self *about* whom it is false.

Just as a direct realist resists the Lockean move that we only directly perceive qualities of physical objects while the object itself remains unseen and unknown, Udayana resists the Humean move that introspection only reveals mental *states*. The internal sense-organ, according to Nyāya, is in direct contact with the self (just as external senses reach out to touch tables and trees). Thus I am acquainted with myself when I am acquainted with a pain or a wish or a cognitive state like my-acquaintance-with-the-table, in me.

This is precisely what Frege argues using two very subtle arguments. First, consider my introspective judgement, "I am looking at the moon." A certain moonward gaze is surely one object of my knowledge. Is not *I*-the gazer another object? If we reduce this I to just another mental state or idea—to a Jamesean *thought* or a Yogācāra *Vijñāna* (an *ālaya vijñāna* or a perverted psychosis—a *kliṣṭamanas*—which imagines itself to be permanent), we face a vicious regress. This mental state or idea will need to be present to consciousness—it has to be *my* idea of myself—in order to come alive. Part of this consciousness will again have to be the pretended-I—the idea of the self. Thus "I should be boxed into myself in this way to infinity."

Secondly, when I am introspectively aware of a lack of experience the self assumes a special claim as an independent object of perception. The content of negative introspective judgements of the very common kind—like "I am not in pain"—cannot be reduced to just a particular perception of a mental event. Even the most ontologically austere analysis of outer perception of absence, e.g., of a pen on a table, cannot get rid of the *locus*—the table as a perceived object. Similarly even if the missed feeling or its lack does not figure in our awareness of painlessness, as the locus of the absence the self *has* to be perceived if the absence has to be felt. This point was anticipated by K. C. Bhattacharya who, like Kant, was unwilling to objectify the self. He explained away such introspection of a want of feeling as my imagining a merely possible pain, or contrasting myself with a third person in pain. But confusing a clear perception of a non-suffering self with the mere non-perception of a suffering one seems to be a sheer mistake.

---

22 Ibid., p. 743.

In Kant also we see a logical link between the self and material objects. But neither the ego nor the spatial permanent of his 'Refutation of Idealism' is transcendentally real. They are both 'made' by understanding. In Nyāya and Frege we have a real self knowing directly both itself and objects which are both independent of experiences and complete (unlike properties which are *incomplete*, i.e., parasitical on objects).

I want to end by reminding you of the fundamental differences between these two realistic schemes which I have more or less bracketed together against anti-realisms of various sorts. Frege's motivation for proving something real over and above subjective states was peculiar: once he had established a self, and then a plurality of selves and objects outside experience, the 'way was clear' for recognizing mind-independent propositions as truth-bearers. These propositions he called "Thoughts" (*Gedanken*). A fact for him is just a true thought. Such thoughts are neither physical nor mental entities. Now, there is no place at all in Nyāya metaphysics for such a third realm. When many of us together perceive an apple to be red, apart from each of us and our unsharable sensations there is only *one* entity according to Nyāya—namely, a red apple out there. Both its substance and its qualities are equally physical. We can analyze the object into a substance, a quality, an action, a universal, a bonding relation of *being in* or exemplification. But on top of these there is no separate proposition or fact *that the apple is red* which I also apprehend. The Fregean tradition continues to speak of true and false thoughts as objective sharable contents of perceptual judgements and as meanings of sentences. Naiyāyikas would be anti-realists about such *propositions*. But that is another story.[23]

[23] A version of this paper was read at the Sixth East West Philosophers' Conference at Hawaii in August 1989. I am thankful to the anonymous referees for this Journal, for suggesting improvements. Much of the research was done during my term (1988-90) as a Jacobsen Fellow at University College London.

Dehātmavāda or the body as soul

EXPLORATION OF A POSSIBILITY WITHIN NYĀYA THOUGHT

PANDIT BADRINATH SHUKLA

According to Nyāya thought, soul is one of the nine substances, the other eight being earth, water, fire, wind, ākāśa, space (*deśa*), time and the mind (*manas*). Soul itself is of two kinds, the *jīvātman* and the *paramātman*. The *jīvātman* is the embodied soul; it is plural, different in every different body.

Soul is distinct from the physical body, the senses, the mind (*manas*) and the life principle (*prāṇa*). It has no form, though it comes into contact with all substances which have a form, and thus becomes an agent and has experiences. Though it has contact with all substances which have a manifest form, yet, due to the power of *adṛṣṭa*, born of earlier *karma*, it experiences joy and suffering only in association with that particular body with which it becomes conjoined due to earlier *karma*. It then engages in new *karma* acquiring sin or merit depending on actions characterized by *dharma* or *adharma*. New experiences create in it new *saṁskāras* (impressions and propensities). In whatever way it acts, the mind joined to it by the force of its *adṛṣṭa* acts as its subordinate.

Such is the embodied soul, the *jīvātman*. It is characterized by nine *guṇas* (qualities/properties) which are specific to it: *buddhi* (cognition, consciousness), desire, aversion, effort, *dharma*, *adharma*, volition, joy and suffering. It is also characterized by five general *guṇas*, namely, number, measure, separateness, conjunction (*saṁyoga*) and disjunction (*vibhāga*), which inhere in it. It is born and reborn in various *yonis* (living forms) according to its *karma*. Only in a human body does it become aware of itself as a candidate (*adhikārin*) for *mokṣa*. In this state it is called *dehī*, *prāṇī* or *jīva*.

The *paramātman* is distinct from the *jīvātmans* or embodied souls; for *paramātman* is one, it is the creater of the world and the author of the Vedas. The *jīvas* who worship *paramātman* are in return bestowed with the gift of endless divine bliss. The *paramātman* impels the *jīvas* to enjoy the fruit of their *karma* and in the attainment of *mokṣa*. It, too, has nine *guṇas* which are its inherent properties: eternal knowledge, desire and effort in addition to number, measure, distinctness, conjunction and disjunction. *Paramātman* is called by names such as Īśvara, Prabhu, Bhagavāna, etc. Since *paramātman*

*The Text of a Special Lecture delivered by Acharya Pandit Badrinath Shukla in Sanskrit at the meeting of pandits organized by the Indian Council of Philosophical Research at Sarnath, Varanasi, from 30 October, 1985 to 2 November, 1985. Translated by Dr Mukund Lath, Department of History, University of Rajasthan, Jaipur.

creates, upholds and destroys the world, he is also given the three names, Brahmā, Viṣṇu and Śiva.

The body cannot have *buddhi* (cognition, consciousness) as an inherent property, because, if this were true, the experience of childhood would not be remembered in old age when the earlier body no longer exists. Moreover, *buddhi* is obviously absent in a dead body. Therefore, clearly, it is not the body that possesses *buddhi*. All other specific *guṇas* of the soul (ātman) such as desire, effort, etc. are rooted in *buddhi* (consciousness, cognition). Therefore, they, too, cannot be inherent properties of the body. Neither can it be said that *buddhi*, desire, effort, etc. are inherent properties, not of the body but of the sense-organs; because the memory of an earlier experience, which was acquired through a sense-organ, persists even after the sense-organ itself is destroyed. Moreover, if each of the sense-organs were separately endowed with *buddhi*, desire, effort and other such *guṇas*. every single body will have a multiplicity of conscious agent (*jñānis*) pulling it in different directions and soon destroying it completely. Further, the mind, according to Nyāya, is atomic, but these *guṇas* such as consciousness, desire, etc. are not; they are thus experienced as continuous in nature. Such an experience would not be possible, if it was the mind and not the *ātman* or self which had these *guṇas* as inherent properties. Therefore, we are led to the conclusion that the *ātman* or self is a substance different from the body, the senses, and the *manas* (mind) and possesses *guṇas* such as knowledge, desire, etc. as inherent properties.

Such is the notion of the self as propounded in the standard texts of Nyāya and the Vaiśeṣika. But it cannot withstand criticism. For it can be demonstrated that within the Nyāya framework the concept of body along with that of the mind (*manas*) can effectively replace the concept of a distinct self, revealing it to be a redundant idea. This new Nyāya doctrine can be called the doctrine of *dehātmavāda*.

According to *dehātmavāda*, the concepts of *manas* and body can successfully fulfil the various purposes for which a distinct self or *ātman* has been posited. And this can be done without giving up the Nyāya framework. The living human body can replace the self as the ground in which those *guṇas* or properties, which are capable of being directly perceived, inhere. These *guṇas* are: *buddhi* (consciousness, cognition), desire, joy, suffering, revulsion and volition. The other three *guṇas* of the self, namely, *dharma* (merit) *adharma* (demerit) and *bhāvanā*[1], which are not capable of being directly perceived, can be taken as inhering in the *manas* (mind). Now it is true that in the Nyāya framework we cannot conceive of the above group of *guṇas*—*buddhi* (consciousness, cognition), desire, etc.—as specific or *viśeṣa guṇas* of the body; for such *guṇas* can arise in the body, an earthly object, only through a process such as *pāka* (maturation, evolution), a process which can result only in giving rise to newer *guṇas* that are essentially similar in nature to the earlier earthly (material) *guṇas*. The limbs of a human body are '*jaḍa*', dead material objects; they have no consciousness, and thus they cannot give rise

to radically different *guṇas* such as *buddhi*, the essence of which is consciousness. But, though we cannot conceive of *buddhi* and like *guṇas* as *specific guṇas* of the body, we can certainly conceive of them as *general* or *sāmānya guṇas* inhering in the body. This is, indeed, our move.

The *general guṇas* in the Nyāya system are such that for them to be conceived of as properties inhering in the body it is not necessary that they be emergent properties, through the process of *pāka*, of the same matter of which the body is constituted. Nor is it necessary that they be similar in character to those essentially material *guṇas* that belong to the various limbs of the body. For this reason it is possible within Nyāya to conceive of *guṇas* like *buddhi* (consciousness, cognition) as belonging to the body and as inhering in it. Just as form, according to Nyāya, inheres in a body as a general *guṇa*, so can *buddhi* and the like. The Naiyāyika cannot but allow this. The body, after all, has other *guṇas* too in his scheme such as *saṁyoga* (contact) and *viyoga* (the state of being disjoined), which inhere in it as *general guṇas*. Similarly, why can we not conceive of *buddhi* and such other *guṇas* as inhering in the body as *general guṇas*?

A question may be posed here: if *buddhi*, desire, effort and other such *guṇas* are conceived of as *guṇas* residing in a body, then why not conceive of them as *specific guṇas* rather than *general guṇas*? This, however, is patently not possible, for this will not be consistent with the very definition of a *specific guṇa*. A *specific guṇa* is defined as a *guṇa* on the basis of which one *dravya* (substance) is distinguished from another. The specificity of a *specific guṇa* has both a positive and a negative aspect. A *specific guṇa* is, by definition, present in substances which are characterized by it; but at the same time it is absent from all other substances. The *guṇa śabdatva* (sound as a property), for example, inheres in *ākāśa* as its *specific guṇa*. It is absent from all other *dravyas*, and is at the same time the *specific guṇa* which distinguishes *ākāśa* from other *dravyas*.

What we are proposing is that *guṇas* such as *buddhi* be conceived of as *general guṇas* inhering in a body as long as it lasts. They are analogous to form (*rūpa*) which, too, in Nyāya is conceived of as a body's *general guṇa*. There are other *guṇas* too, namely, conjunction (*saṁyoga*) and disjunction (*viyoga*), which in Nyāya are conceived of as inhering in the body as its *general guṇas*. *Buddhi* and the like can be similarly conceived of as inhering in a body as its *general guṇas* as long as the body endures as a living thing.

Regarding the three *guṇas*—*dharma*, *adharma* and *bhāvanā*—*dehātmavāda* holds that these inhere in the *manas*, and further that all three are *specific guṇas* of the *manas* since they distinguish *manas* from other substances.

OBJECTION. If *buddhi* and like *guṇas* are conceived of as *guṇas* inhering in a body on the analogy of *rūpa* (form), then they, too, should be directly perceptible like *rūpa*.

ANSWER. There is no rule which stipulates that any single specific sense-organ should be able to perceive all the *guṇas* inherent in a body. We find

that *rūpa* (form) is perceived by the eye, touch by the skin, smell by the nose, and taste by the tongue. Let *manas*, we say, be the organ that perceives *buddhi*. There can be no objection to this.

However, another objection can be raised here, namely, *guṇas* belonging to the body are so conceived in the Nyāya framework that they can be perceived only by outer sense-organs and not by an inner organ like the *manas*. This, indeed, is a rule. Therefore a *guṇa* like *buddhi*, which is to be grasped or perceived only by the *manas*, cannot be a *guṇa* of the body. In reply to this objection, we declare that once we have accepted *buddhi* and other such *guṇas* as 'belonging to' the body, we can certainly further stipulate that *manas*, which is the organ for grasping or perceiving these *guṇas*, can also be taken to be an 'outer' sense-organ, since it is able to perceive *guṇas* such as *buddhi* which belong to the body. The fact that *manas* is called an inner organ is merely indicative of its 'residing within' the body. Moreover, the Nyāya rule is that an outer organ is needed to perceive a *guṇa* that resides on the outside of a body. Since *guṇas* such as *buddhi* reside within the body and not on its surface, for perceiving such *guṇas* what is needed is, indeed, an internal organ such as the *manas*.

Let me explain further. The Sāṁkhya scheme divides sense-organs into two categories: (*i*) *jñānendriyas* (organs of perception) and (*ii*) *karmendriyas* (organs of action). The *manas* (mind) acts as an aid to both these categories of *indriyas*, and is thus called both a *jñānendriya* and a *karmendriya*. On this analogy, taking *buddhi* and other such *guṇas* as belonging to the body, *manas* can be conceived of as an outer and an inner sense-organ: outer because it perceives *guṇas* residing on the outside of the substances, and inner because it resides within the body and perceives 'inner' *guṇas* such as *buddhi*.

OBJECTION. If *buddhi* and other like *guṇas* are conceived of as inhering in the body like *rūpa* (form), then it should be possible for an observer to perceive them just as he can perceive the *rūpa* (form) of another's body.

This, we answer, is absurd. We can never directly perceive *buddhi* and other such *guṇas* belonging to any body. The shape and form of a body can be perceived by another, because they come into the field of the eye's vision. *Buddhi* and other such qualities reside within; hence they never come into the field of an observer's outer vision. Direct perception can result only when an object comes into the range of a sense-organ.

Another objection to our hypothesis can be this: if we assume that it is the body that possesses *buddhi*, then memory would become impossible; an old man with his old body will never be able to remember what he experienced when he was young, for he then possessed a different, young body.

This objection can be easily answered. Our theory of *dehātmavāda* holds that the experiences of a body give rise to *saṁskāras* (impressions) in the *manas*, and that *manas* continues to be the same in the young body as well as the old. It is through the *saṁskāra* residing in the *manas* that an old man can remember what he experienced when he was young. In our theory of

dehātmavāda, experience is conceived of as givingrise to *saṁskāra* which, in turn, gives rise to memory. Experience causes *saṁskāra* in the *manas* through the relation known as *svāśraya-vijātīya-saṁyoga-sambandha:* that is the relation of contact between two distinct and dissimilar objects residing in the same receptacle (*āśraya*). *Saṁskāra* then gives rise to memory in a similar manner. But the relation of the *manas* with the sense-organs or with the limbs of the body is not of the same nature as the relation between the *manas* and the body as a whole. Therefore, neither *saṁskāra* nor memory arise in the sense-organs or in the limbs of a body.

After death, the *manas* associated with the present body enters a new body which is born of the *adṛṣṭa* associated with the present body. This is possible, because the association of the *manas* with the body is a *vijātīya* association. It is for this reason that a newly born baby retains a 'memory' of its *saṁskāras* (impressions) in an earlier body, and begins suckling its mother's breasts as soon as it is born. Its experience in the earlier body had given rise to the *saṁskāra* that suckling the mother's breast is beneficial; and, therefore, in its new birth, too, the 'memory' born of the past *saṁskāra* causes it to suckle its mother's breasts.

OBJECTION. If the experience of a previous body can give rise to memory in a new and different body, then it should also be possible for the experience of one man to give rise to memory in another: what Caitra has experienced should, in this view, be remembered by Maitra.

This objection is groundless, because in the case of Caitra and Maitra the *manas* is not *identical*, whereas in the case of one body being reborn as another the *manas* continues to be the same.

OBJECTION. The present body which you have also equated with the soul engages in actions throughout its life. It cannot, however, attain the fruits of all its actions within the span of a single life time. Therefore, when it dies, some of its actions cannot but be conceived of as destroyed without giving rise to any fruit. How can the *dehātmavādin* fail to arrive at this unseemly conclusion? Further, a new body-soul begins experiencing joys and suffering from the moment of its birth. The *dehātmavādin* cannot account for this. For him these new experiences of joy and suffering must remain fruits of actions never performed.

Our answer is that the *manas* remains common to both the new and the old body. Actions performed in the older body which have not yet borne fruit reside as *saṁskāras* born of *dharma* and *adharma* (merit and demerit) in the *manas*. In the new body born of *adṛṣṭa* (created through *dharma* and *adharma*) in the older body, the same *manas* continues, and thus actions which were performed in the earlier body are enabled to bear fruit in the newer body. Our theory of *dehātmavāda* does not believe in the rule that the fruit of an action is experienced by the same person who performed the action. A belief in this rule is possible only for those who believe in a soul apart from and distinct from the body. This is an old prejudice that we must give up, though,

admittedly, it is difficult to give it up. The rule in which we *dehātmavādins* believe is this: the fruit of an action performed by a body is experienced by another body in which the same *manas* which resided in the earlier body resides. Body and mind (*manas*) in our view are two distinct entities, conjoined with each other through a contact of the *vijātīya-saṁyoga* kind in which two objects can remain in touch without losing their distinct identity.

OBJECTION. It is the body which performs permissible and impermissible actions. If such actions have no interaction with the *manas*, how can they give rise to merit or demerit in the *manas*? And if merit and demerit reside in the *manas* which is quite distinct from the body, how can a new body experience the fruits of earlier action, since the continuity of the new birth with the old is through the *manas* and not through the body?

This objection, we say, is not tenable. We hold that the actions performed by the body cause *adṛṣṭa* in the *manas* which resides in the body through the relation known as *vijātīya-saṁyoga-sambandha*. And then this *adṛṣṭa* which resides in the *manas* becomes the cause of joy and suffering in another body through the same relation.

A further objection may be raised here: the demand for economy of thought would tend to favour positing an independent self or *ātman*, because this would do away with the positing of an indirect causal relation leading from experience to *saṁskāra* to memory on the one hand, and from action to *adṛṣṭa* on the other.

The answer to this objection is as follows: when we choose between two alternative causal explanations, the principle of economy is not by itself sufficient to lead us to the right choice. The totality of what is to be explained should be the prime consideration. The question of economy of thought usually arises in respect to the *form* of definitions where the nature of what is being defined itself is not in question, i.e. all parties agree as to what it is that is being defined, and the choice is to be made only between different formulations of how it is to be characterized in words.[2] Such is not the case in the present situation. The question we have before us concerns the very basic issue as to whether experience, *saṁskāra* and memory can at all be directly related through a causal connection with *karma* and the fruits of *adṛṣṭa*. An appeal to economy of thought cannot be a relevant argument in deciding this issue. Moreover, even if we accept that experience, *saṁskāra*, memory and *adṛṣṭa* reside in the same receptacle, i.e. *ātman*, and thus they can be causally related in a direct manner without necessitating two distinct causal connections, then, too, we shall not really gain in economy; for then we will have also to accept an infinity of all-pervasive, *vibhu*, substances, namely, the *jīvātmans*, and this will lead to another kind of non-economy in thought. This we choose to avoid.

Further, the belief in *ātman* as an extra entity creates other problems. Sentences like 'I go', 'I know' necessitate a basic distinction in the analysis of the two verbs, thus resulting in another loss of economy in thought. In

explaining the first sentence 'I go', traditional Nyāya will have to take the verb 'go' as referring to an action; while the verb 'know' in the second sentence has to be taken as pertaining to the *ātman* in which knowledge resides. *Dehātmavāda* gets rid of this dichotomy. For, if we conceive of the *ātman* as indistinct from the body, both the above sentences can pertain to the same entity. Take also another pair of sentences such as 'Caitra goes' and 'Caitra knows'; the belief in *ātman* as a separate and distinct entity creates a problem in understanding these two sentences. The first sentence 'Caitra goes' may be easily construed as follows: the word 'Caitra' can be understood as referring to Caitra's body which also may be seen as the *āśraya* (ground) of the act of going. But a similar analysis cannot be made of the sentence 'Caitra knows' if we accept the *ātman* theory; because the verb 'know' will then pertain to an entity, i.e. *ātman*, which is distinct from Caitra's body. The two verbs 'go' and 'know' will thus have separate *āśrayas*. Neither can we make the move of taking the verb 'know' as pertaining not to any *āśraya* or entity that knows but to a limited piece of knowledge itself. For, then, the sentence 'God knows all' will be impossible to construe since God's knowledge is unlimited.

However, the following may be postulated by the separate *ātman* theory: the meaning of sentences like 'Caitra goes' and 'Caitra knows' are to be taken as pertaining separately to both the body and the *ātman* of Caitra. But such a move, too, will create a problem. A sentence such as 'Caitra does not know' will then not be able to contradict another statement such as 'Caitra knows', for we will be able to construe the first one as pertaining to the body of Caitra and the second one as pertaining to the *ātman*.

Another objection may be raised against *dehātmavāda* as follows: *dehātmavāda* argues that *saṁskāra* becomes a cause of memory through the fact that both memory and *saṁskāra* reside in the same receptacle (*āśraya*) and are related through a *svāśraya-vijātīya-saṁyoga-sambandha*. It is through this causal connection that memory resides in a human body in the *dehātmavāda* view. But this raises a problem, for a similar causal chain connects memory to the sense-organs too. Hence the sense-organs will also have to be understood as endowed with memory. And, since memory is a kind of knowledge, sense-organs will become identical with *ātman*.

The objection is not justified. We *dehātmavādins* think that the process through which experience gives rise to *saṁskāra* has to be construed in this way. Experience gives rise to *saṁskāra* through the *svāśraya-vijātīya-saṁyoga* relation. And *saṁskāra*, then, gives rise to memory. But this relation *does not* exist between memory and the sense-organs or the limbs of the body. (Memory resides through *manas* in the body *as a whole*, which is an entity distinct from sense-organs and the limbs of the body, which are parts of the body.)[3]

OBJECTION. In *dehātmavāda* the notion of the relation between the *manas* and the body is such that ordinary material objects such as a piece of cloth or a jar can also become associated with the mind (*manas*), and, therefore, they, too, can have *ātman*.[4]

ANSWER. The *manas*, we believe, is related to the body only through an indirect causal connection, i.e. through a causal link established through experience and memory. This is what makes the continuity of *karma* possible for the same *bodymind* entity from one life-time to another. The *manas* has no such causal link with objects like a jar or a piece of cloth. If we make relation such as 'occuring at the same time' or 'being the object of the same knowledge' as equal in status to the relation of *saṁyoga* with the same '*āśraya*' (which the *manas* has with the body) and further argue that the former two relations can also give rise to memory through *saṁskāra*, then the whole world will become the *āśraya* of knowledge and it will have to be believed that everything has an *ātman*. But such an argument is far-fetched and is no more than a vicious attempt to destroy all cogent theorizing. Even in the theory which believes in *ātman* as distinct from the body, not everything can become the *āśraya* of *ātman*. Through relations such as 'occurring together at the same time', the *ātman* can become related to objects such as jars. But this theory denies that knowledge can rise in the *ātman* through its connection with a jar. It is only through the connection of *ātman* with living bodies (which become its *āśraya* through a *vijātīya-saṁyoga*) that the rise of knowledge is possible. Similarly, *dehātmavāda* also believes that, though *manas* can have *saṁyoga* with every thing through relations such as 'occurring at the same time', yet such relations do not give rise to memory or other conscious entities. It is only when *manas* which is the *āśraya* of *saṁskāra* has a *saṁyoga* with a living body that such a *vijātīya-saṁyoga* can give rise to memory, etc. These are matters which can be very easily understood, and to cast unnecessary doubt upon them is misplaced.

But another, a more serious, objection can be brought against *dehātmavāda* by someone who argues as follows: the attempt at repudiating *ātman* and replacing it by the body and the *manas*, in effect, elevates these two to the status of the *ātman*; it does not negate the *ātman* as such.

This argument, too, is not tenable. In our theory the body is non-eternal, whereas the *manas* is eternal. If both together were to form the *ātman*, we shall have to conceive the *ātman* as having two contradictory qualities of being both eternal and non-eternal. This could give rise to ideas contrary to experience, ideas such as 'sometimes I am eternal, but sometimes I am not'. The equation of the pair, body-and-*manas* with *ātman*, is thus not tenable.

Gautama in his *Nyāyasūtra* says: 'Desire, revulsion effort, joy, suffering and *buddhi*, these are what characterizes the *ātman* (*ātmano liṅgam*).' We have accepted all these characteristics as belonging to the body alone and not the *manas*. *Manas* in our postulation is the *āśraya* only of *dharma*, *adharma* and *bhāvanā*. The function of the body and *manas* being so distinct, they cannot be equated with the *ātman* in any sense.

Here, however, is another objection: *manas*, in Nyâya, is atomic. It cannot, therefore, pervade the whole body. How then is consciousness felt to pervade the whole body? The only answer can be to accept an *ātman* which does

pervade the whole body. But this the *dehātmavādin* refuses to do. Thus, his theory cannot explain the common experience of our being able to feel the body as a whole.

OUR REPLY. The theory which conceives *ātman* as distinct from the body also has no answer to the problem. *Manas* in traditional Nyāya remains in contact with the *ātman*, but *manas* can be present only at one tiny part of the body at a single moment. And yet, by the multiplication of these moments, it gives rise in the *buddhi* to experiences that cover the whole body. The *dehātmavādin's* solution to the problem can be similar.

OBJECTION. It is a common human intuition that the *manas* is an internal organ, and also that *manas* is entirely instrumental in character. Its instrumentality is revealed by feelings such as 'Now I am doing this with my mind', 'I am aware through my mind', and the like. *Atman*, however, is not an instrument but is considered to be an agent. *Manas*, being purely instrumental, cannot, therefore, replace it.

This objection is again easily answered. *Dehātmavāda* believes that the *body* is the *ātman* . As for *manas*, it is merely an instrument of this body-as-*ātman*. This we have already stated earlier.

A fresh objection might still arise. If what makes the body conscious is the *adṛṣṭa* which belongs to the *manas*, then it becomes difficult to see how a dead body must be necessarily devoid of consciousness; because, according to *dehātmavāda*, the *manas* containing the *adṛṣṭa* which imparts consciousness to the body continues to exist even after the death of the body with which it was associated.

This objection, too, is not tenable, the reason is that we believe in the rule (*niyaṁa*) that the *vijātīya-saṁyoga-saṁbandha* (the contact between two categorically different objects such as *manas* and the body which makes it possible for the *manas* to be associated with the body) is destroyed immediately and necessarily at the death of the body.

Yet, the following questions may arise: if the body *is* the *ātman*, then usages like 'my body' will have to be understood in a purely metaphorical sense. But in that case, how do we explain the fact that usages such as 'I am the body' are never to be found? How can the *dehātmavādin* explain this?

ANSWER. Linguistic usages depend on our knowledge of both words and the objects they refer to. Since we never have a knowledge which can be expressed as 'I am the body', such usages are not found.

But this only raises a further question: if the body is identical with the *ātman* how then can one explain the fact that such a knowledge never arises?

This question has an easy solution. The meaning of the word 'I' can be grasped only in connection with the characteristics (*guṇas*) of which 'I' can be an appropriate *āśraya*. Therefore, the knowledge of 'I' arises only in terms of 'I am fat', 'I am thin', 'I am happy', 'I am willing', etc. The knowledge such as 'I am the body' does not arise in normal experience; but, then, neither does the knowledge such as 'I am the *ātman*'. The theory, which upholds

ātman as a separate, distinct entity, has, therefore, the problem of explaining why a knowledge such as 'I am *ātman*' is not a common human knowledge. *Ātman* in common experience is known not directly but through its properties such as *buddhi*, joy, etc. resulting in usages such as 'I know', 'I am happy' and the like.

Another objection to *dehātmavāda* can be made on the basis of Patañjali's *Yogasūtras* which speak of various *bhūmis* (aspects or stages) of the *chitta* (psyche). These *bhūmis* such as *kṣipta, vikṣipta* and others have no relation at all with the body or any limb of the body, a fact which is a matter of common experience. *Dehātmavāda* cannot account for their existence. Only the postulation of a distinct *ātman* can do so.

This objection is baseless. *Dehātmavāda*, too, can successfully comprehend and accommodate *bhūmis* of the psyche.

No one has any doubt concerning his own existence. Such doubts as 'Do I exist or Do I not?' just do not arise in anyone's mind. The doubt that does arise is whether the body is the *ātman* or not. For both *dehātmavāda* and the doctrine of a separate *ātman*, the doubt 'Do I exist or not?' will not arise. Since in the *ātman* doctrine the body is decidedly not *ātman* and in *dehātmavāda* it is decidedly so, the question 'Do I exist or not?' is meaningless and adventitious in both cases, and the belief in the existence of self either as the *ātman* or the body remains unquestioned.

What we intend to point out is that the term 'I' refers to the person who utters it. In the *ātman* theory, 'I' will refer to the *ātman* as an entity distinct from the body, implying that the person who says 'I' is an *ātman* distinct from the body. In *dehātmavāda* the same 'I' will refer to the body of the person who utters the personal pronoun. In both cases, 'I' will have a meaningful, unquestionable reference. The doubt 'Do I exist or not?' will in either case be adventitious.

OBJECTION. A statement such as 'He is reflecting on the question whether he is the body or not' will sound very strange if we accept the doctrine of *dehātmavāda*.

This, however, cannot be taken as a serious objection against *dehātmavāda*. In the doctrine of a separate *ātman*, the proposition 'I am not the body' is an unquestionable given. In *dehātmavāda*, on the other hand, what is given as unquestionable is the proposition 'I am the body'. Thus, a question such as 'Am I the body or not?' is adventitious not only for the doctrine of *dehātmavāda* but also for the *ātman* doctrine.

Another objection to *dehātmavāda* can be as follows: the experience that 'I am' seems sometimes to arise from the head and sometimes from the nerves or the flesh of the body. This in *dehātmavāda* is bound to give rise to absurd experiences such as 'I am my head, or 'I am my flesh' or 'I am my nerve'.

Such an objection can only be called crude. Experiences that arise from different limbs of the body such as the head or the flesh or the nerves actually

belong to the body as a whole, which is distinct from its parts and is the real reference of the term 'I'.

YET ANOTHER OBJECTION. The 'I' experience, as we can all feel, is quite distinct from bodily experiences of pain or joy. Therefore, the 'I' experience must be grounded in something, which is quite distinct from both the body and the *manas*.

This objection is again not tenable. We do not believe in the possibility of any experience, which may be characterized as the experience of the pure 'I'; neither do those who believe in the *ātman* doctrine. For both, the meaning of 'I' refers to the person who utters the word. According to us *dehātmavādins*, this person is no different from the body, which is the actual referent of the term 'I'. We do not understand why one should unnecessarily look for another referent.

OBJECTION. In certain states of consciousness such as dreaming, the existence of outer objects including one's own body can become either doubtful, hazy or even controverted. But such a veil of doubt or negation never falls upon the existence of the *ātman*. If the body were the *ātman*, then such an experience should have been impossible in the case of the body too.

We have an answer to this objection. What happens in the above cases is not different from what happens in cases of *bhrama* (illusion), when an object is not perceived in its true character. In a dream the true character of the body as *ātman* becomes veiled by doubt. But this does not mean that we begin to perceive the body as a non-*ātman*, something which it is not, and doubt its truth in the capacity of a non-*ātman*. Such a doubt is not possible.

We believe that the body itself is the *ātman*; there is no *ātman* distinct from the body. Yet we also grant that the body as the ground of actions and efforts is different from the body as the ground of consciousness and the like *guṇas*. In states of dream-like illusion, the perception that we have is *not* that the body is actually a different entity, namely, the *ātman*. The body is, in fact, still taken as the body. What becomes doubtful is the existence of the body as the body, not as something mistaken for the *ātman*, distinct from it.

ANOTHER OBJECTION. The *dehātmavādin* cannot but accept that the final goal of life (*parama puruṣārtha*) is the achievement of physical comfort and material happiness. Yet, we see that human beings are prepared to undergo personal sufferings for the good of others. How can this be explained in *dehātmavāda*?

ANSWER. The doctrine which believes in a distinct *ātman* also has a similar problem, because in that doctrine, too, human action is conceived of as being solely directed towards the attainment of one's own happiness and in getting rid of whatever causes unhappiness. In truth, only a few altruistic persons give up their own happiness and devote themselves to performing actions that would lead to the happiness of others. Such people will continue to exist whether we believe in *dehātmavāda* or in the *ātman* doctrine. There are men who, though they believe in the *ātman* doctrine, are yet ready to act

for the good of others giving up their own personal comforts and accepting pain in the process. Similarly, the *dehātmavādin*, too, if he is a man of sympathy, culture and discernment will devote himself to furthering the happiness of other body-souls, giving up his own happiness and accepting pain in the process. The community of *dehātmavādins* is, in fact, substantial, and among them we do find people who gladly use their wealth for the good of others, opening schools, hospitals and other such philanthropic institutions.

Some thinkers might raise a new objection. In the doctrine of *dehātmavāda*, *adṛṣṭa* and *bhāvanā* are said to belong to *manas*. Now, during the state of *suṣupti* (dreamless sleep) *manas* enters the organ called *purītat* (an organ near the heart) which is mentioned in the Upaniṣads. If this is true, then it will be impossible to explain how the body still keeps breathing during *suṣupti*. To account for this one must accept the existence of a distinct *ātman* which causes the body to keep breathing during that state, an *ātman* which is also the *āśraya* of *adṛṣṭa* and *bhāvanā*.

The objection has no real strength. We believe that the contact between *manas* and the surface of the body (*tvak*) causes consciousness (*jñāna*) which is the basis of other conscious *guṇas* such as desire, revulsion and the like. During *suṣupti* consciousness becomes dormant; and, therefore, desire, revulsion and such other *guṇas* also remain dormant. However, actions such as breathing, which are responsible for maintaining life in the body, do not depend upon consciousness. They depend on *adṛṣṭa* which does not become dormant. Even when the *manas* enters the physical organ called *purītat, adṛṣṭa* actively keeps up such movements in the body which are responsible for breathing as well as other such movements that are the basis of life.

ANSWER OBJECTION. The *dehātmavādin* cannot really explain all of man's actions in terms of their fruits. The actions performed by a man towards the end of his life do not give rise to results during the life time; and, therefore, such actions are bound to remain fruitless and thus meaningless if we accept the doctrine of *dehātmavāda*. Why should a man, then, engage in such actions?

ANSWER. The *dehātmavādin* believes that the fruit of a man's actions need not accrue to him alone but can accrue to others who survive him. In this manner, actions performed by a man towards the end of his life can also have their fruit. It is wrong to say that man acts only for his own good. He also acts for the good of others as is, indeed, clear from the actions of men. It cannot be said that those men, who perform actions aiming at the good of others, do so with the purpose that, if their actions are not fruitful during their own lifetime then the merit (*puṇya*) resulting from them will yield them fruit in subsequent lives (*janmāntara*). For it is seen that people, who believe that this life is all that we have and that there is no *janmāntara*, yet engage in good deeds throughout their life, the results of which are enjoyed by others.

FRESH OBJECTION. There is another argument that can establish the existence of *ātman* as distinct from the body. The argument is as follows: 'The body being an assemblage of parts is meant for the sake of another like a bed

which is a similar assemblage.' The existence of the body, according to this argument, establishes the existence of *ātman* for whose sake it is assembled. This argument cannot be answered by a mere battery of words or by subtle casuistry. Yet, we do have an answer. The objection is, in fact, neither clear nor cogent. The notion 'for the sake of another' (*parārtha*) is not a clear notion. If 'for the sake of another' means 'for the enjoyment of another distinct from itself', then it is difficult to see how the argument can prove the existence of a separate *ātman* through the example of the bed. For, even if a bed is meant for another, it is difficult to see why this 'another' should be the *ātman*. We can take this 'another' to be the body. The *ātman* doctrine, however, cannot agree to this interpretation, since it does not believe that a body can be an enjoyer. But, then, if 'for the sake of another' is taken to mean 'that which does not itself enjoy but is meant for the enjoyment of another', then, too, the argument will fail. It will not serve the purpose of the *ātman* doctrine, for it will fail to apply to the body. Because (as we believe) the body as a whole is distinct from a mere aggregate of its parts, it will not be proper to call it a mere 'assemblage'. If, in order to save the argument, we modify our argument and say 'the body is for the sake of another, for it is a created object', then, too, the argument will remain unconvincing. Any created object, which is meant 'for the sake of another', has to be a *jaḍa* object, something made up of dead matter; but the body, though admittedly a created object, is not a *jaḍa* object, and is thus not 'for the sake of another'.

ANOTHER OBJECTION. *Dehātmavāda* makes activities such as performing Vedic sacrifices pointless.

ANSWER. This is not really true. Firstly, because in our doctrine sacrifices such as *putreṣṭi*, which aim at bearing fruit in this very life, do retain a purpose. Secondly, sacrifices which are said to result in the attainment of *svarga* can also be meaningfully performed by a *dehātmavadin,* because *svarga* is said to be an object desirable for everyone; and so a *dehātmavādin,* too, can desire it and so perform sacrifices that aim at its attainment. However, it may yet be said that, according to *dehātmavāda, svarga* cannot really be attained since it is not attainable by a body. This is certainly true, but it does not constitute a major objection. Firstly, because sacrifices may be performed for the enhancement of one's prestige, if not for *svarga*; secondly, results of sacrifices which aim at a mundane fruit can be attainable by a body which may not always be the present body, but will still be the home of the same transmigrating *manas* in another life. Many sacrifices, moreover, are meant for the benefit of others; *dehātmavāda* quite approves of these, because, as we have said earlier, it is human nature to engage in actions which result in the good of others.

A FURTHER OBJECTION. It is not really possible to conceive of *punarjanma* (transmigration) in the *dehātmavāda* scheme.

Our answer to this is that, even in the doctrine of a distinct *ātman, punarjanma* is impossible to conceive of, for it presents the same problems of identity as it does in *dehātmavāda*. If all we mean by *punarjanma* is that the same

ātman comes into contact with a new body through a *vijātīya-saṁyoga* relation, thus maintaining identity through different lives, then such an identity in which the *same* person is said to be reborn is conceivable in *dehātmavāda* too, because we believe that the *manas* continues to exist after the death of the body. *Manas*, in our view, is the *āśraya* of *adṛṣṭa* through which it acquires contact with a new body and is thus reborn. The theory of *punarjanma* can, in this sense, be upheld even within our framework.

OBJECTION. The doctrine of *bandha* and *mokṣa* (of being fettered to *saṁsāra* and of liberation in *mokṣa*) becomes meaningless in *dehātmavāda*.

OUR ANSWER: *Bandha* is just another name for engaging in actions which cause *adṛṣṭa*. The *adṛṣṭa*, then, leads to results which can only be experienced in a new life through a new body. Such a conception of *bandha* is quite tenable in *dehātmavāda* too. And *mokṣa* after, all is nothing but the absence of *bandha*. We believe that a body which has not realized its own body-soul nature through yoga should be called *baddha* (fettered to the world of transmigration); for such a body continues to perform actions which result in *adṛṣṭa* leading to fruits that have to be enjoyed in a new life. But a man who has realized his body-soul nature does not engage in such actions, and is thus 'free' or 'liberated'.

FURTHER OBJECTION. *Dehātmavāda*, in fact, cannot avoid the view that after death both *baddha* (bound) and free persons are really reduced to naught without a trace; so there is no real difference between being *baddha* and being free. Why should, then, any 'body-soul' strive for the realization of truth, giving up the pursuit of palpable sensory pleasures?

Such an objection, we must say, can be brought against the theory of a distinct soul also. For, in that doctrine too, the liberated soul is no different from being totally dead or extinguished (*mṛtopama*).

OBJECTION. There appears to be no real point in positing the new doctrine of *dehātmavāda*. For all that this doctrine has to say is that an ever-continuing (*nitya*) *manas* keeps transmigrating from one body to another, bearing *adṛṣṭa* and *saṁskāra* acquired through experiences in an earlier body; that the new bodies into which this *manas* transmigrates serve merely as vehicles for remembering experiences of the older bodies and for experiencing the results of actions done through them.

ANSWER. I am sure that this much will be generally granted that our position is an improvement in terms of economy of thought on the traditional Nyāya-Vaiśeṣika doctrine, which posits an infinite numbers of all pervading (*vibhu*) souls. The doctrine of distinct and separate *ātman* has also much else that is cumbersome resulting in an unnecessary *gaurava* (multiplication of entities and relations) in thought. It, first, posits an endless array of all-pervading souls, and then is forced to conceive of infinite relations over infinite moments with infinite substances and forms into which these souls enter. Our doctrine avoids such cumbersomeness.

Moreover, the doctrine of a separate *ātman* cannot avoid taking an amoral

stance regarding human action. Since, in that view, men are determined totally by their previous *karmas* and their *adṛṣṭa*, they are powerless against exploitation and tyranny. The *ātman* doctrine does not permit men to do anything about such things, for they are not free to do so. They are not free to remove inequalities from any given social and economic set-up, nor can they fight against a cruel government indifferent to the welfare of its subjects. *Dehatmavāda* is open to the idea that new action can be undertaken by a new body-soul. No earlier *karma* is powerful enough to constrain a man to acquiesce passively in the exploitation of one man by another under the belief that this is an inevitable result of earlier *karma*. The community of *dehātmavādins* is free to engage in actions aiming at changing the present conditions and creating a more just social and economic order beneficial to them all. They are free to create a more beautiful world.

OBJECTION. This is mere wishful thinking, for *dehātmavāda* will actually encourage people to seek their own selfish ends without caring for others. Self-seeking is a common human failing; and if one is not made responsible for one's actions beyond death, then there will be no reason for a man to desist from seeking his own selfish ends without caring for the suffering and exploitation of others.

Such considerations, however, need not antagonize us towards *dehātmavāda*. The moving spirits behind selfless actions are great selfless men of the past. The prestige that is attached to their great deeds aimed at the common good, and the reverence shown to them in history books should be enough to give rise to a similar impulse in others.

Another commendable thing about *dehātmavāda* is that it can influence people to improve themselves in this very life, since improvement in an afterlife is not possible. Listening to the great tales of great men, a *dehātmavādin* will be moved to try and improve himself in this very life. In the *ātman* doctrine, the temptation of postponing a good action and leaving it for another life is very strong. A man is more likely to pursue mean and selfish ends under that scheme than under *dehātmavāda*. *Dehātmavāda* is, consequently, not only more rational but also more moral.

TRANSLATOR'S NOTES

My translation is an attempt to present to the philosophically inclined English reader a non-technical version of Badrinathji's Sanskrit essay. Badrinathji was a philosopher of great originality—as this essay, I think, also evinces—but being a Navya-Naiyāyika, he assumed a knowledge of Navya-Nyāya technical vocabulary in his readers. This was natural enough, since not only Navya-Nyāya but a great deal of intellectual writing in Sanskrit assumes such a knowledge. Most disciplines in Sanskrit that touch upon philosophy—and few do not—have been using Navya-Nyāya vocabulary and techniques for the sake of a clearer articulation of concepts.

I have not tried to translate these technicalities. Attempting a closer technical translation of Badrinathji's essay would have presented hurdles which we are not yet quite able to cross. There is no satisfactory standard English version of Navya-Nyāya vocabulary and

modes of expression. And even if we were to have one, it would need years of scholarly and creative cultivation before it can become really entrenched. As it is, those who can read Navya-Nyāya in some kind of translationese, can also understand Sanskrit. Such people will discover that my English version is lacking in certain other ways, too. I have, for example, not translated a quotation or two from Navya-Nyāya texts which Badrinathji's original includes. Badrinathji assumes a close and easy familiarity with the texts he quotes. To the English reader, not familiar with the Navya-Nyāya *paramparā*, the quotations, I think, would have sounded merely scholastic, and redundant at that. Badrinathji, moreover, has no footnotes. This is a modern habit, but sometimes useful. I have resorted to it at one place (fn. 2) where I felt that what Badrinathji had to say was intended to be in a kind of parenthesis.

On the whole, however, I have tried to remain as close to the original as possible, following the steps of the argument as it moves.

Notes and References

1. *Bhāvanā* in *Nyāya* is another name for *saṁskāra*, a property of the soul that makes it possible for experience to leave its impression or traces upon the soul. The term *saṁskāra*, however, has a larger application; it applies not only to conscious *jīvas*, but also to 'dead' matter—*vega* (speed) thus is a *saṁskāra* of wind (*vāyu*). It is through *bhāvanā* that memory becomes possible. *Bhāvanā*, in turn, cannot be directly perceived, but only inferred from the fact of memory.
2. Let us take an example. The Mīmāṁsakas make the following analysis of the process of inference. Inference, they say, is a result of two discrete cognitions: (1) *Sādhyavyāpyo hetuḥ* (the *hetu*—that through which one wishes to prove, the middle term, is pervaded by the *Sādhya*—the 'cause' of the *hetu*, the major term); and (2) *hetumān pakṣaḥ* (the *pakṣa*—the locus—possesses the *hetu*). The Naiyāyikas, who disagree with the Mīmāṁsakas concerning the proper analysis of the process of inference, yet agree with them that these cognitions do arise. But they argue that it is unnecessary to accept two separate cognitions in order to characterize correctly the process of how inference is caused. A single cognition, they say, will do, namely, *Sādhyavyāpyahetumān pakṣaḥ* (the *pakṣa* possess the *hetu* which is pervaded by the *śādhya*). The Mīmāṁsaka has no quarrel with the Naiyāyika concerning the fact that such a cognition does occur; he differs as to its relevance to a proper analysis of how inference arises.

   The Naiyāyika analysis, we say, is to be preferred, because it has the virtue of 'economy' (*lāghava*). Naiyāyikas further argue that their acceptance of a single cognition as the cause of inference has another virtue. Recognizing two *separate* cognitions as necessary for the rise of inference can result in a problem. Inferences arise in human beings. If we grant the necessity of two separate cognitions for it to arise, then our analysis will not be able to negate cases where two *different* persons might each have *one* of these cognitions. In order to avoid this difficulty, the *Mīmāṁsaka* might make the move of inserting a further stipulation in his analysis, namely, 'The two cognitions necessary for inference to rise must be possessed by the same person.' This will only result in further cumbrousness in his analysis. For it will then become necessary to make a *separate* causal analysis for each different case of the occurrence of the *same* inference.

   A similar problem occurs in the analysis of *śābdabodha* (understanding language). All Naiyāyikas agree that for *śābdabodha* to occur a knowledge of *yogyatā* is a necessary condition. *Yogyatā* is a kind of existential constraint and must be observed in using language: thus a usage such as 'wets with fire' lacks *yogyatā*, for 'wetting' and 'with fire' do not, in fact, go together, and this *fact* renders the sentence meaningless. Any philosopher defining *śābdabodha* must be careful to include the knowledge of *yogyatā* as one of the necessary factors within the body of the definition itself. Other-

wise the same kind of difficulty that we spoke of earlier in connection with defining how inference arises will pose a hurdle: the knowledge of *yogyatā* in one man, will not be able in our definition to prevent the rise of *śābdabodha* in another. The definition will become cumbersome.

3. The parenthesis is to make the point clear. It is not part of the original.
4. This objection perhaps needs a clarification. In the *dehātmavāda* view, the body is a material substance and *guṇas* such as *buddhi,* desire, effort and the like, inhere in it as *general guṇas.* Badrinathji likens these *general guṇas* to *guṇas* such as form (*rūpa*). The *manas* becomes associated with *buddhi,* desire, effort and such *general guṇas* indirectly, through its association with the body. The *general guṇas* of the body are destroyed with the body. The *manas* which is a category apart, is not destroyed and becomes attached to another body. The problem with this view which Badrinathji anticipates in this objection can be stated in terms of two related questions. One, since the *general guṇas* are so conceived that any material substance can have them, how is it that only a human body has them? And two, since the *manas* becomes associated with the *general guṇas* only through its association with a material substance, why is it that such an association takes place only in the human body and not in other material substances such as a jar or a piece of cloth.

This page intentionally left blank

Gerald James Larson An eccentric ghost in the machine: Formal and quantitative aspects of the Sāṁkhya-Yoga dualism

Within the "grammar" of dualisms in the history of philosophy, the classical Sāṁkhya position appears to be something of an anomaly. The Sāṁkhya is a bit like the *nitya-samāsa* in Sanskrit composition, that is to say, a compound that cannot be analyzed according to the conventional rules. By this I mean that the Sāṁkhya does not fit the usual or conventional notions of dualism. If one looks, for example, at the classic expression of the dualist position in Western thought, namely that of Descartes, one realizes immediately that the Sāṁkhya somehow misses the mark. In his *Principles of Philosophy* Descartes comments as follows about the dualist problem:

Extension in length, breadth and depth constitutes the nature of corporeal substance; and thought the nature of thinking substance. For every other thing that can be attributed to body presupposes extension and is only some mode of an extended thing; as all the properties we discover in the mind are only diverse modes of thinking (cited in K. Nielsen, Reason and Practice, 1971, p. 332, I: liii).

In his *Meditations* Descartes sets forth the essence of the dualist perspective as follows:

Because on the one hand I have a clear and distinct idea of myself, in as far as I am only a thinking and unextended thing, and so on the other hand I possess a distinct idea of body, in as far as it is only an extended and unthinking thing, it is certain that I, that is my mind by which I am what I am, is entirely and truly distinct from my body and may exist without it (Nielsen, VI).

A more recent statement of the conventional dualist position is that of the analytic philosopher Kai Nielsen, who puts the matter as follows:

The core of the dualist claim . . . could . . . be put in this way: There are at least two radically different kinds of reality, existence or phenomena: the physical and the mental. . . . Physical phenomena or realities are extended in space and are perceptually public or, like electrons and photons, are constituents of things that are perceptually public. . . . Mental phenomena or realities, by contrast, are unextended, not in space, and are *inherently* private (Nielsen, p. 333).

Whether one considers the Cartesian position or, according to Kai Nielsen, the modern, analytic restatement of it, the interpreter of Sāṁkhya must admit that the Sāṁkhya is *not* a dualism in these senses. Similarly, if one considers the theological or ethical dualism of Christian thought—*à la* Pauline theology or later treatments such as those of Augustine, and so forth—again, the interpreter of Sāṁkhya must say that Sāṁkhya is not a dualism in these senses. Similarly, if one considers the dualistic analyses in Plato or Aristotle, or the Kantian dualism of noumena and phenomena, or a phenomenological dualism of *noesis* and *noema*, the Sāṁkhya is not really dualist in any of these senses. Even within the framework of Indian philosophy, the garden-variety dualisms of the later

Gerald James Larson is Professor of Religious Studies at the University of California, Santa Barbara.
AUTHOR'S NOTE: *This article was prepared for a Symposium on Sāṁkhya and Yoga at the Institute for the Study of World Religions, SUNY, Stony Brook, New York, April 1981.*

Vedānta schools or the older archaic *jīva-ajīva* dualism of the Jains does not adequately fit the Sāṁkhya case.

Indeed, vis-à-vis all of these positions, I am inclined to argue that Sāṁkhya represents a critique of the traditional or conventional dualist position and approaches, rather, the opposite position, or what modern Western philosophy of mind would call "reductive materialism," that is to say, a philosophical view which "reduces" "mind"-talk or "mentalistic"-talk to "brain-process"-talk or, in other words, construes mind, thought, ideas, sensations, and so forth, in terms of some sort of material stuff, or energy, or force (as has been argued, for example, by H. Feigl, J. J. C. Smart, Kai Nielsen, and others). For, according to classical Sāṁkhya, the experiences of intellect (*buddhi*), ego (*ahaṁkāra*), and mind (*manas*), and the "raw-feels" such as "pain" (*duḥkha*) or "pleasure" (*sukha*)—or, in other words, what conventional dualists would consider to be "inherently private"—are simply subtle reflections of a primordial materiality (*prakṛti*)—a primordial materiality undergoing continuous transformation (*pariṇāma*) via its constituent modalities of externalizing activity (*rajas*), reflexive discriminating (*sattva*), and reifying objectivation (*tamas*). Thus, the modern reductive materialist's claim that "sensations are identical with certain brain processes" would have a peculiar counterpart in the classical Sāṁkhya claim that "awarenesses" (*citta-vṛtti*s or *antaḥkaraṇa-vṛtti*s) are identical with certain *guṇa*-modalities. Or, again, the modern reductive materialist's claim that the conventional notions of "the inherently private" or "the mental" are only linguistic fictions which inhibit a more correct understanding of the human situation would find its peculiar counterpart in the classical Sāṁkhya claim that the notion of the discreet "individual" or the "individual ego" seriously inhibits a more correct understanding of the basically "dividual" and "transactional" environment in which "human existence" occurs (to use McKim Marriott's idiom). Both positions, in other words, appear to criticize the notion of an inherently private, mentalistic "ghost in the machine" as being a product of verbal carelessness (*vikalpa*) brought about by the failure to make relevant distinctions (*aviveka* or *avidyā*).

Alas, however, the comparison of Sāṁkhya philosophy with reductive materialism quickly breaks down, for instead of expelling the traditional or conventional "ghost in the machine" and getting on with the task of describing the world without "ghost"-talk, Sāṁkhya, as it were, refurbishes the "ghost," stripping it of its conventional attributes and reintroducing it as what I am calling in this paper "an eccentric ghost," eccentric in the sense that it no longer has anything to do with "mind"-talk or "mentalist"-talk or "ego"-talk, all of which latter are fully reducible to *guṇa*-talk in good reductive materialist fashion. Sāṁkhya designates its "eccentric ghost" as "consciousness" (*puruṣa*), thus necessitating the differentiation of "awareness" (*citta-vṛtti* or *antaḥkaraṇa-vṛtti*) from "consciousness" (*puruṣa*) and requiring a radically different kind of dualism, namely, a dualism between a closed causal system of reductive materi-

alism (encompassing "awareness" or the "private" life of the mind), on the one hand, and a nonintentional and contentless "consciousness," on the other. It thereby rejects idealism without giving up an ultimately transcendent consciousness. It rejects traditional or conventional dualism by reducing "mentalist"-talk to one or another transformation of material "awareness"; and it corrects reductive materialism by introducing an "eccentric ghost in the machine" which is nonintentional and has nothing to do with ordinary mental awareness. It is a classic case of a philosophy that wants to have its idealist/realist/materialist cake and eat it too! Whether or not Sāṁkhya philosophy is finally successful, of course, is open to debate and interpretation—the Indian tradition itself, it should be noted, took a rather dim view of it—but I would argue that it raises some interesting philosophical issues that may still be relevant in current discussions within the philosophy of mind.

As is well-known, it is difficult to trace the origins of this peculiar Sāṁkhya dualism and its "eccentric ghost in the machine", since detailed texts that lay out the basic arguments and methodology of the system are simply not available from any period. What remains are summaries and digests of the system (for example, the *Sāṁkhyakārikā*, the *Tattvasamāsa-sūtra*, and the very late *Sāṁkhya-sūtra*, together with a variety of commentaries the great majority of which are generally useless). Also, there are various attacks on Sāṁkhya by various other schools of Indian philosophizing, some of which can be used to reconstruct some of the details of the system; yet the highly polemical environment in which these discussions occur can hardly reassure the student of Sāṁkhya that he or she is getting a fair account of the system. K. C. Bhattacharya has expressed the matter well:

> Much of Sāṁkhya literature appears to have been lost, and there seems to be no continuity of tradition from ancient times up to the age of the commentators.... The interpretation of all ancient systems requires a constructive effort; but while in the case of some systems where we have a large volume of literature and a continuity of tradition, the construction is mainly of the nature of translation of ideas into modern concepts, here in Sāṁkhya the construction at many places involves supplying of missing links from one's imagination. It is risky work, but unless one does it one cannot be said to understand Sāṁkhya as a philosophy. It is a task that one is obliged to undertake. It is a fascinating task because Sāṁkhya is a bold constructive philosophy (*Studies in Philosophy*, 1956, I: 127).

Recently in my own attempts at the "supplying of missing links" I have been exploring various formal or quantitative "paradigms" in the interpretation of classical Sāṁkhya, mainly because the term *sāṁkhya* itself appears to be related to the notion of "number" or "enumeration" and because the extant texts of the system present elaborate enumerations in what appears to be a paradigmatic fashion. The problem, of course, is one of identifying the sorts of paradigms that the ancient Sāṁkhya *ācāryas* had in mind. Unfortunately, as Karl H. Potter and others have observed, Indian philosophy generally (including Sāṁkhya philos-

ophy) makes no clear distinction between analytic and synthetic statements or a priori and a posteriori judgments as in classical and modern Western logic, and, therefore, whatever formal or quantitative paradigms there are in Indian philosophy cannot be directly related to what one might consider to be an obvious source for formal paradigms, namely, traditions of Indian logic. To be sure, there are some fascinating relations between Sāṁkhya and ancient Indian logic, and Randle has suggested that the earliest reflections in India on the nature of inference can be traced to Vaiśeṣika and Sāṁkhya philosophy. Nevertheless, the Sāṁkhya enumerations are not primarily intelligible in terms of the standard categories or paradigms of Indian logic, or, putting the matter another way, Sāṁkhya philosophy does not appear to be directly relevant to the standard issues and problems of Indian logic. The basic thrust of the system would appear to be in quite a different direction.

Formalism, however, is not solely or perhaps even primarily the turf of the field of logic, as modern structuralism has taught us. Such divergent fields as physics, mathematics, music, psychology, anthropology, and linguistics have all benefitted and advanced via systems theory, semiotics, structural analysis, and other varieties of formal analysis, and it is to be noted that ancient India made some interesting contributions in most of these fields. Indeed, one could easily mount an argument that India's most important contributions to knowledge were not really in the areas of logic or technical philosophy but, rather, in such areas as grammar, phonology, mathematics, psychology, and anthropology.

In any case, in an attempt at the "supplying of missing links" for the interpretation of classical Sāṁkhya and, specifically, for the interpretation of Sāṁkhya's "eccentric ghost in the machine" (namely, *puruṣa*), I want to call attention to three distinct (yet related) paradigmatic approaches to the Sāṁkhya enumerations, namely,

(a) a "mathematical" paradigm;
(b) a "linguistic (phonological and grammatical)" paradigm; and
(c) a "dyadic" or "contrastive feature" paradigm.

By "paradigmatic approaches" I mean metaphors of modelling that may have been used by the Sāṁkhya *ācārya*s as they pursued their speculative philosophizing. I am not suggesting, in other words, that Sāṁkhya is a form of mathematics, or a form of grammar, or a form of structuralist or "dyadic" analysis. I am suggesting, rather, that the ancient Sāṁkhya *ācārya*s may have used patterns or models such as these as heuristic metaphors for organizing their thought. (It is conceivable, of course, that the Sāṁkhya *ācārya*s were themselves personally unaware that paradigms or models such as these related to their style of thinking. That would not disprove the validity of such paradigmatic approaches, however, any more than a native speaker's ability to speak a language fluently and correctly is related to the speaker's knowledge of the formal grammar of the language. That is to say, a native speaker may speak quite correctly without any knowledge of the grammar.)

(*a*) *The "mathematical" paradigm.* All traditions of Indian learning proceed by making lists of things—for example, in astronomy, phonology, grammar, medicine, prosody, philosophy, logic, mythology, meditation, and so forth. The student of ancient Indian learning soon grows weary of this particular cultural predilection—a predilection typical, to be sure, of many ancient cultures but perhaps carried to an extreme in India. Ancient India was obsessed with numbering. Even in love-making (*Kama-sūtra*) there is a set number of permissible postures, a set number of aphrodisiacs, a set number of useful flirtatious gestures, and a set number of techniques for arousing a lover. To a large extent, of course, enumerations are useful devices for preserving oral traditions of learning, and undoubtedly most enumerations are quite arbitrary and conventional mnemonic devices. Should anyone doubt this, I would simply invite the doubter to read any of the dreadful Abhidhamma books of the Buddhist Pali Canon—an exercise in reading very much like reading the Santa Barbara or Stony Brook telephone directory!

In some instances, however, patterns of enumeration become more than arbitrary mnemonic devices. In Pāṇini, for example, rules are enumerated in a precise order so that the sequence of numbering is related in important ways to the application of grammatical rules. Or again, as E. G. McClain and A. T. de Nicolás have argued, obscure Vedic numberings appear to be related to certain common ratios in musical acoustics (when reduced to simple whole numbers). Moreover, certain numbers appear again and again—for example, certain square numbers such as 25 (the Sāṁkhya *tattvas*), 36 (Kaśmir Śaivism), 49 (the letters of the Sanskrit "alphabet"), 64 and 81 (the number of Yoginīs in Indian iconography), and so forth. Interestingly also—as was pointed out to me by Thomas Hopkins of Franklin and Marshall College—if one multiplies 1^1 by 2^2 by 3^3 one gets $1 \times 4 \times 27$ or "108," one of the most common "symbolic" numbers in Indian literature.

Regarding the Sāṁkhya system, for many years I had ignored the specific enumerations and had more or less followed along with the traditional view that the enumerations were only arbitrary mnemonic devices. While working with the *Tattvasamāsa-sūtra*, however, in recent years—a text which contains only twenty-five short numerical utterances—I began to wonder if the pattern of enumerations had additional significance beyond that of mnemonic convenience, an additional significance perhaps on analogy with Pāṇini's pattern of enumeration. I therefore began to play with the Sāṁkhya numbers, reducing larger numbers to smaller numbers (and operating with the methodological principle that only those numbers that actually appear prominently in actual Sāṁkhya texts could be used in the "game"). What finally emerged, interestingly, was not related to Pāṇini's sequences but was related, rather, to a phenomenon in the field of mathematics. That is to say, there is a clear preference or predilection in Sāṁkhya philosophy for the sequence of prime numbers (numbers whose factors are only themselves or 1), namely, 2, 3, 5, 7, 11, 13, 17,

19, 23, and so forth. Moreover, between 1 and 100 there are exactly twenty-five prime numbers, and, of course, according to classical Sāṁkhya philosophy, there are twenty-five basic *tattvas* or fundamental principles. All numbers presuppose number 1, and the number 1, therefore, must perforce be the *mūlaprakṛti*; and the sequence of Sāṁkhya "entities" can be exhibited as follows:

(1) *mūlaprakṛti* (and implicit in *prakṛti* are the three *guṇas* that manifest themselves in pairings of 2 + 1);
(2) *buddhi* (with its twofold *bhāva*-structure);
(3) *ahaṁkāra* (with its threefold structure: *vaikṛta*, *bhūtādi*, *taijasa*);
(5) *tanmātras* (five subtle elements)
(7) five *tanmātras* + *buddhi* + *ahaṁkāra*, referred to as "the seven" in *Sāṁkhyakārikā* III;
(11) *indriyas*, including the five sense capacities, the five action capacities, and *manas* or "mind";
(13) *liṅga* or *karaṇa*, the thirteenfold instrument made up of *buddhi*, *ahaṁkāra manas*, and the ten sense capacities and which transmigrates from life to life;
(17) structure of *ahaṁkāra* when fully manifest (namely the elevenfold *vaikṛta-ahaṁkāra*, the fivefold *bhūtādi-ahaṁkāra*, and the one *taijasa-ahaṁkāra*);
(19) transmigrating entity empowered by *prakṛti* (namely, the thirteenfold instrument + the five *tanmātras* + *prakṛti*, according to *Sāṁkhyakārikā* XLII; or, in its Vedānta variant as set forth in Śaṅkara's *Bhāṣya* to the *Māṇḍūkya Upaniṣad*: "nineteen mouths," namely, the five sense capacities, the five action capacities, the five *prāṇas* + *manas* + *ahaṁkāra* + *buddhi* + *citta*; compare *Bhāṣya* in verse 2 of the *Māṇḍūkya*);
(23) the manifest world that emerges because of the co-presence of *prakṛti* and *puruṣa* (from "Brahmā down to a blade of grass," according to *Sāṁkhyakārikā* LIV).

The twenty-fifth *tattva*, of course, is *puruṣa*, and the Sāṁkhya texts describe the *puruṣa* as contentless consciousness whose presence allows *prakṛti* to become manifest. Given the number sequence that has emerged, there is, of course, only one possibility for *puruṣa*, namely, zero (0), a notion which was not only known to the ancient Hindus but possibly discovered by them. The notion of zero is necessary for all sophisticated calculation, yet it has the peculiar characteristic of not adding anything in any calculation. It is an irreducible principle necessary in any sophisticated theory of numbers, yet it is not clear, even in modern mathematics, if zero itself can be construed as a number.

From such a perspective, Sāṁkhya philosophy can be construed as generating the natural world utilizing a "mathematical" model or paradigm in a manner not unlike that of ancient Pythagorean philosophy. The number 1 is the source of all whole numbers, the primes being particularly important in that their factors are only themselves and 1, with composite numbers being reducible to the primes. Such a perspective would suggest that the primes 2, 3, 5, 7, and so forth, could serve as building blocks, as it were, in a "numbered world," and this, of course, they appear to do in Sāṁkhya philosophy. One might also usefully compare the

generation of the Sāṁkhya *tanmātras* and *mahābhūtas* (namely the so-called "accumulation theory") to the Pythagorean notion of the *gnomons* (the so-called "carpenter's squares") by means of which the sequence 1, 2, 3, 4, 5 generates the squares 1, 4, 9, 16, and 25. The perfect number 10 of the Pythagoreans also appears to have its counterpart in the 10 *cūlikārthas* or *mūlikārthas* of the Sāṁkhya. Implicit in such a "mathematical" approach to Sāṁkhya, of course, would be the corollary notion that the notion of "number" links up in some intelligible way with the notion of "subtle matter" or "thing" or "manifest entity," and here again a comparison with Pythagoreanism is relevant, for Pythagoreans wanted to correlate "numbers" with "things."

In any case, construing Sāṁkhya as a kind of archaic mathematical physics (on analogy with Pythagoreanism) may provide one useful avenue for attempting to decipher the nature of the peculiar Sāṁkhya dualism and its "eccentric ghost in the machine." (I have published portions of the foregoing analysis in an article entitled "The Format of Technical Philosophical Writing in Ancient India: Inadequacies of Conventional Translations," *Philosophy East and West* 30, no. 3 (July 1980): 375–380.)

(*b*) *The "linguistic" paradigm.* If ancient Indian traditions of learning were obsessed with numbers and enumerations, they were also almost as equally obsessed with language and grammar, and here again the paradigms or models of linguistic reflection provide some useful perspectives in interpreting the peculiar Sāṁkhya dualism and its "eccentric ghost in the machine." Some years ago (in an article entitled "A Possible Mystical Interpretation of *ahaṁkara* and the *tanmātras* in the Sāṁkhya," in the *Sri Aurobindo* commemoration volume ((Pondicherry, 1972), pp. 79–87) I noted some striking similarities between Sāṁkhya philosophy and ancient phonological treatises (*śikṣā*, *prātiśākhya* literature, and so forth). I observed, for example, the appearance of Sāṁkhya-like terminology in phonological texts (for example, *prakṛti*, *vikāra*, *vyakta*, *saṁyoga*, *karaṇa guṇa*, *mātrā*, and so forth). I was especially interested in the terms *ahaṁkāra* and *tanmātra*, and I related *ahaṁkāra* to the notion of *Oṁkāra* ("pronouncing or reciting Oṁ") and *tanmātra* to the notion of *mātrā*, a term referring to the length of time required for pronouncing a short vowel. A short vowel is one *mātrā*; a long vowel two *mātrās*; an elongated vowel three *mātrās*; a consonant or stop is $\frac{1}{2}$ *mātrā*; $\frac{1}{4}$ *mātrā* is an *aṇu* ("moment" or "atom"); and $\frac{1}{2}$ *aṇu* is a *paramāṇu*. (See *Taittirīya Prātiśākhya* as discussed by W. H. Allen in his *Phonetics in Ancient India*.) As is perhaps obvious, I then related *ahaṁkāra* and the *tanmātras* to Oṁkāra and its *mātrās* as set forth in the little *Māṇḍūkya Upaniṣad* and other texts, arguing that *ahaṁkāra* and the *tanmātras* in Sāṁkhya seem to have roots in the old Upaniṣads and in environments where speculation centered around *mantras*, mystical syllables, and mystical sounds. Creation takes place by means of naming, speaking, uttering sounds. The *mūlaprakṛti* would then be potential sound or unmanifest sound. The manifestations of

prakṛti would be, on one level, various series of sounds and combinations of sounds. Also, the method of achieving *kaivalya*, described in *Sāṁkhyakārikā* LXIV as *tattvābhyāsa*, may echo a similar environment. *Abhyāsa* means "study," "analysis," and so forth, but it also means "repetition" or "repeated recitation." By repeatedly reciting the *tattvas* one gradually comes to know the basis of creation of emergence itself, the creative pattern of sounds. The *puruṣa* principle would then be that which transcends all sound—passive, empty silence which is yet the background and presupposition of all sound distinctions.

Recently I have been attempting to take such a "linguistic" paradigm a bit further into the area of grammar and speech itself, and I have found some further useful perspectives for thinking about the Sāṁkhya dualism and the "eccentric ghost in the machine." As is well-known, Sāṁkhya asserts that intellect (*buddhi*), ego (*ahaṁkāra*), and mind (*manas*)—in other words, *antaḥkaraṇa-vṛtti* or *citta-vṛtti*—are manifestations of *guṇapariṇāma* (the "transformations" or activity of *sattva*, *rajas*, and *tamas*) but have nothing whatever to do with *puruṣa*. Consciousness (*puruṣa*) is simply "present" (*sākṣitva*) to *antaḥkaraṇa-vṛtti* or *citta-vṛtti*, and the only interaction is one of mutual reflection. Consciousness is described as being wholly inactive (*akartṛbhāva*), and it is described as being wholly uninvolved in the transactions of *antaḥkaraṇa-vṛtti* (or, in other words, the transactions of *buddhi*, *ahaṁkāra*, and *manas*). All interpreters have agreed (not only modern Western scholars but ancient Indian thinkers as well) that this is a peculiar way of thinking about consciousness, and, more than that, that it appears to be rather absurd on the face of it. It may not be so absurd, however, if the Sāṁkhya *ācāryas* were thinking about the basic *tattvas* from the perspective of a certain grammatical paradigm or model having to do with the basic structure of nouns and verbs in a sentence. The paradigm or model I have in mind relates to nominal inflection in Sanskrit in which the Sanskrit grammarians make a distinction between what they call *vibhakti* (case endings) and *kāraka* (case relationships). There are eight cases in Sanskrit (nominative, accusative, instrumental, dative, ablative, genitive, locative, and vocative). The case endings or case terminations are as follows:

prathamā—"first"—nominative
dvitīya—"second"—accusative
tṛtiya—"third"—instrumental
caturthī—"fourth"—dative
pañcamī—"fifth"—ablative
ṣaṣṭhī—"sixth"—genitive
saptamī—"seventh"—locative
"eighth"—vocative (although most grammarians denied that the "eighth" is really a "case")

These case endings, or terminations, however, must be clearly distinguished from what the grammarians called "case relationships" or *kārakas*. Although there are seven case terminations (or eight, if one counts the vocative), there are only

six essential case relationships—a case relationship being the manner in which the nouns and adjectives in a sentence relate to the verbal action of the sentence. The six essential case relationships or *kāraka*s are as follows:

> *kartṛ*—nominative or subject of the action
> *karman*—accusative or object of the action
> *karaṇa*—instrumental or instrument or means of the action
> *sampradāna*—dative or recipient of the action
> *apādāna*—ablative or ablation of the action
> *adhikaraṇa*—locative or location of the action

The *kāraka* expresses what a sentence "does" or how it acts or functions. What is to be said, then, about the so-called "sixth case" or genitive? Interestingly, according to the Sanskrit grammarians, it has no *kāraka*—it is in no way related to the verbal action of the sentence. The sixth case or genitive only relates one *thing* to another, and it is wholly uninvolved in the action of the sentence. It is called simply the *sambandha* or the case of mere relation. Extrapolating from this grammatical paradigm, the Sāṁkhya *ācārya*s may have been thinking about *buddhi*, *ahaṁkara*, and *manas* as being involved in the transactions of *prakṛti*—the "verbal action," as it were, of *guṇapariṇāma*—and thereby having *kāraka*-functions within the transformations of the manifest world. Consciousness or *puruṣa*, however, has nothing whatever to do with the transactions of *prakṛti* and is only present (*sākṣitva*) to *prakṛti* as mere relation. An intriguing piece of evidence for this kind of linguistic paradigm or model is the old Sāṁkhya claim that inference has to do with certain kinds of real relations between entities. In the *Jayamaṅgalā*, for example, reference is made to an old Sāṁkhya interpretation of "inference" as relating to "seven basic relations" (*sapta-sambandha*):

(1) *sva-svāmibhāva-sambandha*—possessor and possessed
for example, a king and his subject
(2) *prakṛti-vikāra-sambandha*—principal and secondary
for example, barley and groats
(3) *kārya-kārana-sambandha*—effect and cause (material)
for example, cow and calf
(?) (4) *pātra-pātrika-sambandha*—measure and measured
for example, stick and what it measures (perhaps whole and part)
(5) *sāhacarya-sambandha*—companionship or association
for example, a pair of *cakravāka* birds
(6) *pratidvandvi-sambandha*—opposition or hostility
for example, cold and heat
(7) *nimitta-naimittika-sambandha*—efficient cause and effect
for example, eater and eaten (*Jayamaṅgalā* on Kārika V).

Dignāga, Kamalaśīla, and other Buddhist philosophers also make reference to an old Sāṁkhya notion of "seven basic relations" (*sapta-sambandha*), and their listings of the basic relations, though somewhat different in terms of examples, are basically in keeping with the above list from the *Jayamaṅgalā*. Of these seven relations, only the first is typical of the relation between *puruṣa* and *prakṛti*; the other six are all internal relations within *prakṛti* (either in terms of the analytic

distinction between "whole" and "part" or in terms of the synthetic distinction between "cause" and "effect"). Quite apart from the sophistication or lack thereof in this kind of characterization of inference (and it must be admitted that Indian philosophy soon moved beyond the discussion of inference in terms of real relations), the Sāṁkhya discussions of the *sapta-sambandha* indicate that the ancient *ācāryas* may have been philosophizing with one eye, as it were, on models of thinking derived from traditions of grammatical categories and paradigms. Consciousness or *puruṣa*, then, is that entity which has no relation to the verbal action of the world but is in a relation of mere presence very much like that exhibited by the sixth or genitive case.

(*c*) *The "dyadic" or "contrastive feature" paradigm.* The third possibly productive approach to understanding the peculiar Sāṁkhya dualism and its "eccentric ghost in the machine" is to look at the basic Sāṁkhya inferences (set forth in the first twenty or so of the *kārikās*) from the perspective of a kind of "structuralist" analysis. Here I am not so much referring to a model or paradigm used as a metaphor as I am to what appears to be the actual methodology of Sāṁkhya thinking. It clearly resonates to the paradigms to which I have already referred (namely the "mathematical" and the "linguistic"), but there appears to be no source for it other than the Sāṁkhya texts themselves. Possibly it is a unique Sāṁkhya approach and represents Sāṁkhya's peculiar contribution to India's philosophical heritage.

The actual contents of the Sāṁkhya inferences (which are mainly of the *sāmānyato-dṛṣṭa* type) are reasonably clear and may be found described in any number of textbooks on Indian philosophy. What is perhaps not so clear is how the various inferences and basic Sāṁkhya notions relate to one another. Unfortunately, again, there is precious little information in the texts and commentaries regarding this problem, and one must proceed by reconstructing possible scenarios from what is actually available. The only test for such reconstructions is whether or not they render the system as a whole more coherent.

What is most striking to the careful reader of the Sāṁkhya texts is the tendency towards dyadic or contrastive analysis. Dyads, polarities, oppositions, and contraries appear everywhere, so much so that it becomes clear, as already indicated, that dyadic or contrastive feature analysis represents a kind of overall methodology in the system. Some of the more important of these dyads are the following:

(1) *avyakta/vyakta* (unmanifest/manifest)
(2) *prakṛti/vikṛti* (creative/created)
(3) *sāttvika buddhi/tāmasa buddhi* (discriminating intellect/reifying intellect)
(4) *dharma/adharma* (virtue/lack of virtue)
(5) *jñāna/ajñāna* (knowledge/lack of knowledge)
(6) *vairāgya/avairāgya* (dispassion/lack of dispassion)
(7) *aiśvarya/anaiśvarya* (power/lack of power)

(8) *sāttvika ahaṁkāra/tāmasa ahaṁkāra* (discriminating egoity/reifying egoity)
(9) *antaḥkaraṇa/bāhyakaraṇa* (internal organ/external organ)
(10) *apavarga/bhoga* (discriminating experience/ordinary experience)
(11) *nirupabhoga/upabhoga* (without experience/ordinary experience)
(12) *liṅga/bhāva* (inherited disposition/projecting disposition)
(13) *sūkṣma/sthūla* (subtle/gross)
(14) *sūkṣma-śarīra/sthūla-śarīra* (subtle body/gross body)
(15) *aviśeṣa/viśeṣa* (nonspecific/specific)

Moreover, even the triad of the *guṇas* is basically dyadic in structure inasmuch as the *guṇas* always relate with one another in ratios of two-plus-one, or, putting it in the framework of the manifest structure of reality, there is an ongoing polarity between *sāttvika* and *tāmasa* manifestations with *rajas* playing largely a mediating role ("... *taijasād ubhayam*," Kārikā 25). *Sāttvika* manifestations become the reflexive, discriminating "awarenesses" characteristic of the properly functioning *buddhi* when internal or external objects of awareness or activities of awareness have been correctly assimilated. They manifest themselves in ordinary experience as feelings of joyous and quiet fulfillment (*śānta*). *Tāmasa* manifestations are the intentional and reified "objects" of awareness themselves (whether internal in terms of one's own self or ego being an object, or external in terms of subtle elements or gross objects). They manifest themselves in ordinary experience as feelings of alienation and reification (*mūḍha*). The *guṇa*, *rajas*, however, is not a manifestation at all, but is, rather, the externalizing process of life itself (*pravṛtti* or *karman*), which pervades ordinary human existence and generates feelings of discomfort (*ghora*) and a sense or feeling of tragic suffering (*duḥkha*) (or, as the *Yuktidīpikā* puts it: "*duḥkhaṁ raja iti*"). All "private" and "public" manifestations or experiences, therefore, are more adequately talked about simply in terms of the modalities of the *guṇas* (*guṇa-pariṇāma*, or, as the common refrain puts it: *guṇā guṇeṣu vartanta iti*). Then, too, all experiences and manifestations are construed in terms of *satkārya*, or the notion of causality that posits an identity between cause and effect (or *kāraṇa-kārya*). Both analytically and synthetically, therefore, we are dealing with a closed causal system of reductive materialism which can be rewritten in brief as "*guṇā guṇeṣu vartanta iti*." From an analytic point of view, every "component" of the system is a "part" of the totally functioning "whole" (and may well explain why the Sāṁkhya lends itself to a purely mathematical formulation as hinted at earlier). From a synthetic point of view, every empirical manifestation is an "effect" that is finally a mere modification of one ultimate, unconscious (*acetana*) material "cause" (*mūlaprakṛti*).

That the Sāṁkhya system is so tight as a closed causal system of reductive materialism led Debiprasad Chattopadhyaya to suggest in his book *Lokāyata* (third edition, Calcutta, 1973) that originally Sāṁkhya indeed was a pure materialism derivative of an archaic, matriarchal, Tantric milieu. The notion of *puruṣa*, or what I am calling the "eccentric ghost in the machine," was a later

interpolation into the system when Indian intellectual life came to be dominated by patriarchal and largely Vedāntin notions of male consciousness.

Be that as it may, the extant classical Sāṁkhya texts clearly make reference to *puruṣa* or "consciousness" as "something" totally distinct from the closed causal system of reductive materialism. These references occur primarily in the sequence of inferences set forth in Kārikās IX–XXI (and in all of the accompanying commentaries). Again, the analysis unfolds via a series of dyads or contrastive features. In the second hemistich of verse II, after rejecting perceptible and scriptural means for overcoming suffering, Īśvarakṛṣṇa calls attention to a "superior means" for overcoming suffering that differs from both the perceptible and scriptural means, namely, the proper discrimination (*vijñāna*) of *vyakta*, *avyakta*, and *jña*: *vyaktāvyaktajña-vijñānāt*. He then introduces the basic Sāṁkhya dualism in the following fashion (verse III):

mūlaprakṛtir avikṛtir mahadādyāḥ prakṛti-vikṛtayaḥ sapta, ṣoḍaśakas tu vikāro na prakṛtir na vikṛtiḥ puruṣaḥ.

Primordial materiality is uncreated; the seven, intellect, and so forth, are both created and creative. The sixteen are created; consciousness is neither created nor creative.

The four hemistichs of the verse may be exhibited as follows:

(I) *mūlaprakṛtir avikṛtir,*
(II) *mahadādyāḥ prakṛti-vikṛtayaḥ sapta;*
(III) *ṣoḍaśakas tu vikāro,*
(IV) *na prakṛtir na vikṛtiḥ puruṣaḥ.*

Here *puruṣa* is clearly distinguished from all other tattvas in the sense of not being implicated in *prakṛti* or *vikṛti* (and in a formulation, if I might add parenthetically, that is strikingly reminiscent of Johannes Scottus Eriugena's opening passage in Book I of his *Periphyseon*, which reads as follows:

"It is my opinion that the division of Nature by means of four differences results in four species, (being divided) first into that which creates and is not created, secondly into that which is created and also creates, thirdly into that which is created and does not create, while the fourth neither creates nor is created" (Sheldon-Williams edition, Dublin, 1968).

In this Sāṁkhya formulation, moreover, it is to be noted that the first part (namely, *mūlaprakṛtir avikṛtir*) is a negation of the third part (*ṣoḍaśakas tu vikāro*) (and it is a negation of the *paryudāsa*-type); and the fourth part (namely, *na prakṛtir na vikṛtiḥ puruṣaḥ*) is a negation of the second part (*mahadādyāḥ prakṛti-vikṛtayaḥ sapta*). It may be anticipated, therefore, that whatever is predicated of the second part will provide us with a negative description of the fourth part, and whatever is predicated of the third part will provide us with negative descriptions of *both* the first part *and* the fourth part (inasmuch as the fourth part is similar to the first part to the extent that it, too, is uncreated). And,

alas, the descriptions that emerge in Kārikā X and following provide us with precisely these descriptions. The descriptions may be exhibited as follows:

	(1)	(*puruṣa*) —	(*prakṛti*)
Predicates of the third part	(2)	(*jña* — *avyakta*)	(*vyakta*)
		ahetumat (uncaused)	*hetumat* (caused)
		nitya (nontemporal)	*anitya* (temporal)
		vyāpin (nonspatial)	*avyāpin* (spatial)
		asakriya (stable)	*sakriya* (unstable)
		eka (simple)	*aneka* (complex)
		anāśrita (unsupported)	*āśrita* (supported)
		aliṅga (nonmergent)	*liṅga* (mergent)
		anavayava (nonparts)	*avayava* (parts)
		aparatantra (independent)	*paratantra* (contingent)
Predicates of the second part	(3)	(*jña*)	(*avyakta* — vyakta)
		atriguṇas (no *guṇas*)	*triguṇa* (having three *guṇas*)
		vivekin (differentiated)	*avivekin* (undifferentiated)
		a-viṣaya (nonobject)	*viṣaya* (an object)
		asāmānya (uncharacterized)	*sāmānya* (verbally characterizable)
		cetana (conscious)	*acetana* (unconscious)
		aprasavadharmin (unproductive)	*prasavadharmin* (productive)
Predicates from Kārikās XVII–XIX	(4)	*sākṣitva* (witnessing)	
		kaivalya (isolation)	
		mādhyasthya (indifference)	
		draṣṭṛtva (presupposition of experience)	
		akartṛbhāva (inactivity)	
		bhoktṛbhāva (presupposition of subjectivity)	
		bahutva (plurality)	

If the "mathematical" paradigm discussed earlier is reminiscent of Pythagoreanism, then this final "dyadic" or "contrastive feature" paradigm, at least with respect to *puruṣa*, is somewhat reminiscent of medieval "negative theology" and older traditions of Neoplatonism. Hence, the parallel with Johannes Scottus Eriugena's *Periphyseon* may be more than a curious similarity. There may be, in other words, a deeper intellectual affinity, possibly as a result of historical contact in the Hellenistic era or possibly as a result of a common Indo-European cultural heritage. (Regarding possible historical contact in the Hellenistic era, if I may be permitted another parenthetical observation, it would appear that classical Sāṁkhya or older traditions of Sāṁkhya-Yoga can be usefully com-

pared to Neo-Pythagoreanism and Neoplatonism. As is well known, a number of scholars have suggested possible links between ancient Greek reflection and Indian thought, usually arguing for possible links between Greek thought and the ancient Upaniṣads. My own view is that such linkage is quite likely but hardly through the murky ambiguity of Upaniṣadic thought. A more likely point of contact would have been through the ancient traditions of Sāṁkhya and Yoga, both of which traditions were taking shape after the third century B.C. and received technical formulation in the third and fourth centuries A.D. Sāṁkhya is perhaps the oldest tradition of Indian philosophy; it has interesting affinities with early Buddhist thought; it was linked from ancient times with traditions of ascetics and wandering *sādhus*; and, perhaps most important, its early traditions were taking shape precisely at the time when the first contacts with Greek thought could have occurred. The *Sāṁkhyakārikā* of the fourth century purports to be a summary or digest of an older system called the "*ṣaṣṭitantra*," or system of "sixty-topics," and this older system could easily have been in existence between the second century B.C. and the first two centuries A.D. Moreover, Sāṁkhya philosophy, along with another ancient tradition of thought, the Vaiśeṣika, appears to be the theoretical basis of ancient Indian medicine, and we know from the careful work of Jean Filliozat that Indian and Greek medical practitioners were already in contact at the Persian courts from the fifth and fourth centuries B.C. onward. It is quite possible, therefore, that notions of Sāṁkhya and Yoga could have been known in the Mediterranean world of the Hellenistic era, or possibly earlier. Moreover, the point of contact could well have been through traditions of ancient medical speculation, mathematics, and language—subject matters which would have had considerable practical *and* theoretical interest for both sides in any cross-cultural encounter. Unfortunately, there is insufficient evidence to document such encounters. We know only that the vague cultural memories that Neo-Pythagoreans, Plotinus and the Neoplatonists, various Gnostic groups, and others somehow learned were important things from India.)

Returning, however, to our "eccentric ghost in the machine," let me conclude by pulling together what Sāṁkhya appears to be wanting to say about its "eccentric ghost" based on the paradigmatic approaches set forth in this paper. Consciousness or *puruṣa* is a contentless "something," essential to experience and the world but not involved directly in the transactions or manifestations of the world. It is a simple "presence" (*sākṣitva*) in the world, one that enlivens the intellect, ego, and mind but one that has no *kāraka*-function—a mysterious silence that makes speaking possible but is itself unutterable. It is outside the realm of causality, outside of space and time, inactive, utterly simple, unrelated, uninvolved in emergence or transformation, without parts, completely independent, apart from the three *guṇas*, transcendent yet always immanent, the presupposition of discrimination or differentiation, neither an object nor

a subject, unintelligible or verbally uncharacterizable, unproductive, a pure witness whose only relation to *prakṛti* is sheer presence, utterly isolated, completely indifferent, the ground of being, nonagent, and present in the "awareness" of all sentient beings as not being that awareness.

For Sāṁkhya, what is irreducibly important is the simple fact of our presence to ourselves. The authentic mystery in our existence, that before which we must become silent, that in the light of which all wonder begins, and that which enables us to see that there is something rather than nothing, is, according to Sāṁkhya, our presence to ourselves—a presence which is at one and the same time a radical foundation for experience and a radical foundation for freedom.

This page intentionally left blank

Mind/Consciousness Dualism in Sāṅkhya-Yoga Philosophy*

PAUL SCHWEIZER
University of Edinburgh

1 Introduction

In the western philosophical tradition, there is a very prevalent tendency to equate the mind with the conscious self. This tendency is exemplified in the framework of Descartes' *Meditations*, which, not withstanding the abundant criticism this text has received over the centuries, in some sense still provides the most definitive and well articulated expression of the presuppositions underlying the modern western conception of mind. Descartes unhesitatingly identifies the referent of the indexical expression 'I' with *res cogitans*, the thinking substance, and thereby conceives the mind and self as one. For Descartes, thinking substance is of course metaphysically independent of extended substance, which engenders his familiar (and surprisingly resilient) dualism between mind and matter. Also without hesitation, Descartes takes conscious thought and subjective experience to be unqualified constituents of mental substance, and this particular mode of classification has become so much a part of our conceptual heritage that the dividing line he draws (even if not the attendant metaphysical commitment) seems almost self-evident. Surely, if a dualism with matter is to be entertained, then thoughts and subjective presentations belong on the *non-material* side of the dividing line.

In contrast to this fairly ubiquitous Cartesian view, I would like to examine an alternative but intimately related version of substance dualism advanced by two allied schools (*darśana*) of orthodox Hindu philosophy, namely the Sāṅkhya school, and classical Yoga as expounded by Patañjali.[1]

* I would like to thank an anonymous referee of *PPR* for helpful comments.

1 By using the appellation 'Sāṅkhya-Yoga' philosophy, I do not mean to endorse the view that the two schools are essentially the same, nor that Yoga is simply an applied or practical component grafted onto Sāṅkhya theory. Rather, I merely wish to discuss some (of the many) theoretical points which the two schools have in common, though I have tended to use predominantly Sāṅkhya terminology.

These schools draw a metaphysical dividing line not between mind and matter, but rather between matter and consciousness, wherein the mind is placed on the *material* side of the ontological divide. I think that this alternate formulation of the mind/body problem constitutes a valuable intellectual perspective for contemporary western philosophy. It illuminates the problem from a different angle, and this in turn highlights some of the presuppositions underlying the western analysis, and reveals that in an important sense these presuppositions constitute *choices* about how to conceptualize the phenomena, rather than inherent divisions supported by the phenomena themselves. In particular, the Cartesian identification of mind with the conscious subject will be brought into sharp focus, as well as the related Humean identification of the self with a constellation of subjective ideas.

In section 2 of the present paper, I will give a general overview of the Indian position, and in sections 3 and 4 I will explore some of the theoretical consequences of this alternative dualism, and critically compare these consequences with some recent manifestations of the Cartesian model. A primary tenet of the paper is that the classical Indian division between pure consciousness and matter has distinct philosophical interest, especially with respect to current discussion on the relationship between consciousness and mental representation, as well as the endeavor to give a computational and/or naturalistic account of mentality.

The philosophical interest of this alternative dualism is at least partly due to the fact that Descartes' schism between mind and matter has proved extremely adaptable, and reappears under a variety of contemporary guises. For example, Searle (1980) accuses proponents of 'strong' AI (artificial intelligence) of embracing a rigid separation between mind and matter, since they view the mind as a program of rules for symbol manipulation, and this is to treat the mind as an abstract formalism, fundamentally independent of the physical medium in which it happens to be instantiated. Thus certain advocates of the very recent discipline of AI seem committed to a strict separation between the mental and the physical. In a related general vein, the computational and formal treatments of mind advocated by cognitive science and AI are by nature syntactical, even when considered at the level of physical instantiation, and as such they seem to say nothing about the 'real meanings' or semantical contents which are quintessentially mental. Indeed, the deep theoretical gulf between syntax and semantics can be viewed as a highly specialized and refined variation on Descartes' theme. Syntax (at the level of actual tokens) is a system of physical characters, while semantical content seems to be something distinctively *cognitive*, and the connection between the two, at least with respect to computational theories of mind and natural language processing, is a very problematic issue.

Another well known guise in which standard Cartesian dualism finds expression is the infamous gap between qualia, *i.e.*, the 'elements' of subjective conscious experience, and the associated neurophysiological configurations of the brain. Subjective 'ideas' or presentations are held by Descartes to belong to the realm of *res cogitans*, whereas the brain obviously belongs to the realm of *res extensa*. This gap between qualia and brain states constitutes a theoretical focal point of modern dualism, since the correlation between objective physical states of the brain and subjective conscious experience is the most intimate level of correspondence between matter and mind, and it seems to form the crucial boundary line where the two substances should 'intersect.' As I will argue below, the Sāṅkhya-Yoga analysis, though still dualistic, is less problematic with regard to the representational content of conscious mental states, as well as with regard to the functional role of qualia, since content and consciousness are held to be metaphysically independent.

2 Sāṅkhya-Yoga Dualism

The Sāṅkhya school, or *darśana*, is one of the oldest philosophical traditions of India, and many of its ideas are traceable to the Ṛgveda and the early Upaniṣads. Its historical founder is Kapila, though the original *Sāṅkhyasūtras* (aphorisms) he is said to have written during the 6th or 7th century B.C. are now lost, and the most important of the existing texts is the *Sāṅkhyakārikā* (explanatory verses) of Īśvarakṛṣṇa, from around the 3rd century A.D. The Sāṅkhya tradition has a great many theoretical points in common with the classical Yoga *darśana* as expounded in Patañjali's *Yogasūtra*, probably written somewhere from the 4th to the 2nd century B.C.,[2] and the metaphysical position discussed in the present paper is part of their shared philosophical framework. Thus the basic dualism between consciousness and matter, as well as more specific allied points, will henceforth be referred to, somewhat generically, as the 'Sāṅkhya-Yoga view.'

According to the Sāṅkhya-Yoga view, the ultimate principle underlying matter is *prakṛti*, the metaphysical substrate supporting all material phenomena.[3] The mind is deemed to be part of the material world, and hence to be metaphysically grounded in *prakṛti*. The body is also part of the realm of *prakṛti*, and thus mind and body are held to be of the same metaphysical

2 Textual chronology is of course a notoriously difficult and controversial affair. The dates provided in this paragraph are taken from Raju (1985).

3 In the ensuing discussion, the terms '*prakṛti*' and 'matter' will be used roughly as synonyms, even though this is not strictly correct, if matter is construed in terms of the 'particles' which make up physical objects. *Prakṛti* is the *metaphysical* principle which underlies physical manifestations, though for expository convenience I will often equate the physical world with *prakṛti*.

substance. In its most general connotation, the term 'mind' in western philosophy corresponds to the combination or complex of three faculties recognized in the Sāṅkhya-Yoga taxonomy, namely *manas*, *buddhi*, and *ahamkāra*, whose respective significance will be briefly outlined below.

Manas (which is often translated directly as 'mind', though it is only a single facet of the 'mental triplex') is viewed essentially as an organ, the special organ of cognition, just as the eyes are the special organs of sight. Indeed, *manas* is held to be intimately connected with perception, since the raw data supplied by the senses must be ordered and categorized with respect to a conceptual scheme before various objects can be perceived as members of their respective categories, and as inhabiting a world characterized by the systematic and distinguishable attributes normally perceived. This imposition of conceptual structure on the chaotic field of raw sensation is one of the basal activities of *manas*, and forms the distinction between brute sensation (*nirvikalpaka*) as opposed to differentiated perception (*savikalpaka*). Hence ordinary perceptual experience is already heavily conditioned by the activities of *manas*, and *manas* is thus sometimes referred to as the sixth organ of sensation.

In addition to its perceptual activities, *manas* is held to be responsible for the cognitive functions of analysis, deliberation and decision. It is closely allied to *buddhi*, which is somewhat roughly translated as the faculty of 'intellect' or 'reason.' *Buddhi* is a subtler and more powerful faculty than *manas*, and is responsible for the higher level intellectual functions, which require intuition, insight and reflection. The Indian *buddhi* is in some ways comparable to the Greek *nous*, while *manas* is responsible for lower level discursive thought and analysis. But *buddhi* is still regarded as a manifestation of *prakṛti*, albeit the most subtle and refined form which material substance can assume. The combination of *manas* and *buddhi* roughly correspond to what is meant by the objective or 'impersonal' mental faculties in western philosophical discourse. In addition, Sāṅkhya-Yoga recognizes a third component of mind, *ahamkāra*, which is the ego or phenomenal self. *Ahamkāra* appropriates all mental experiences to itself, and thus 'personalizes' the objective activities of *manas* and *buddhi* by assuming possession of them. The combination of these three faculties is referred to as *antahkaraṇa*, the 'inner instrument,' which approximately comprises the individual mind-self of the western philosophical tradition.

But on the Sāṅkhya-Yoga account, the realm of *prakṛti* or matter is held to be inherently unconscious, and is thereby incapable of producing consciousness as an effect. The manifestations of *prakṛti* are always objects, and it is argued that objects can never transform themselves into subjects. Thus at the heart of this dualistic position is the notion that mind-material is not capable of *generating* consciousness out of unconscious ingredients.

Subjective awareness is a distinct ontological category, and in principle it cannot be derived from the stuff of which objects are made. So, in sharp contrast to the western approach, the mind and the cognitive activities it sustains are held to be intrinsically unconscious, since *manas*, *buddhi* and *ahamkāra* are all manifestations of *prakṛti*.

According to the ancient view, the dynamics of *prakṛti* are governed by the interactions of the three *guṇas*, which are the three basic types of constituent of physical substance. The three *guṇas* are *sattva*, *rajas* and *tamas*, which correspond roughly with 'transparency and buoyancy,' 'energy and activity,' and 'inertia and obstruction.' All physical phenomena are believed to consist of unstable mixtures of these three types of constituent, and the instability of these mixtures is responsible for the evolution and transformations of the material world. Thus the conceptual processes sustained by the mind are governed by the mechanical and unconscious interplay of the *guṇas*, and to this extent, mental phenomena are viewed in purely 'physicalistic' or mechanical terms. The unfolding of thought-forms is an integral part of the evolution of *prakṛti*, and mental processes are simply the result of appropriate transformations of unconscious material substance. It is worth noting at this point that the Sāṅkhya-Yoga view thereby avoids one of the most serious pitfalls of Cartesian dualism, since on the Indian account, mental causation does not violate physical conservation laws. By including the mind in the realm of matter, mental events are granted causal efficacy, and are therefore able to directly initiate bodily motions.

Consciousness, on the other hand, is held to belong to a different ontological category altogether. Consciousness is placed in the realm of *puruṣa*, the absolute, unconditioned self, which in some respects is comparable to Kant's noumenal self. *Puruṣa* is described as pure and undifferentiated awareness, and it is held to be immutable and inactive, to be formless and without parts or limiting characteristics. *Puruṣa* is the metaphysical principle underlying the individual person, and closely corresponds to the *ātman* of the Vedānta school.[4] *Puruṣa* is held to exist in complete independence of the material realm, and so the basic dualism in the Sāṅkhya-Yoga metaphysics is between *puruṣa* and *prakṛti*, between consciousness and matter. On this account, mind belongs to the world of matter, while the self belongs to the realm of pure awareness, and thus the mind and the self are ontologically distinct. Movement and form are characteristics of matter, and they are also characteristics of thought, which is a manifestation of matter, while consciousness is held to be intrinsically formless and unchanging.

---

[4] It should be noted, however, that Sankhya-Yoga recognizes a multiplicity of distinct *puruṣas*, i.e. one for each self, while Vedāntic thought tends to view the *ātman* as ultimately singular. Hence on the former account, there is a numerically distinct *puruṣa* associated with each material mind-complex.

What then is the relationship between consciousness and matter, and how are conscious mental events possible? According to the Sāṅkhya-Yoga model, thought processes and mental events are conscious only to the extent that they receive external 'illumination' from *puruṣa*. Consciousness is standardly compared to a light, which illuminates the specific material configurations or 'shapes' assumed by the mind. It is the subtle 'thought-material' of the *buddhi* which allows mental events to appear conscious, because the refined *buddhi* substance is transparent to the light of consciousness. Thus conscious thoughts and perceptual experiences take place when *buddhi* receives representational forms, both perceptual and conceptual, from *manas*, the organ of cognition. *Buddhi* is believed to consist of a preponderance of the 'transparent' *sattva guṇa*, and thus the representational forms it receives from *manas* are capable of becoming translucent with the light of consciousness. So *buddhi* receives cognitive structures from *manas*, and conscious 'light' from *puruṣa*, and in this manner, specific mental structures are capable of being illuminated by an external source, and thereby these structures are able to appear conscious. But consciousness itself is entirely independent of the particular thought structures it happens to illuminate.

The translucent quality of *buddhi* distinguishes thought-stuff from the gross material objects of thought and perception, which ordinarily contain a preponderance of the 'dark' or opaque *tamas guṇa*. Thus only the thought-material of *buddhi* is capable of conscious illumination, rather than the entire material realm, since a preponderance of the *tamas guṇa* renders the external objects of the material realm opaque to the light of consciousness. Hence representational structures, rather than the objects represented, are capable of conscious illumination, which is why minds appear sentient while stones and tables do not. Only the subtle stuff of the mind is a suitable medium for receiving sentience, and thus it is minds which are the loci of awareness in the natural world. In this manner, conscious thought processes and attendant representational content is comparable to photographic slides of external objects. The photographic image stored in the film is composed of matter, but it is both representational and translucent, and therefore is analogous to the thought-structures which glow with the sentience of *puruṣa*

So, on this model, the representational content of thought is carried in the unconscious physical configurations of the mind, and certain of these patterns or configurations become illuminated by an external and undifferentiated awareness, resulting in the phenomenon of particular conscious thoughts and subjective experiences. But consciousness and the illuminated mental processes are entirely independent. Formless and immutable consciousness plays no causal role in the transformation of mental structures, but rather is a passive 'witness' to some small portion of these structural

transformations, *viz.* that portion normally identified with the conscious individual or Cartesian ego. According to the Sāṅkhya-Yoga analysis, the Cartesian or phenomenal ego is not the real self, but rather is a partial reflection of the self in the determinate structures of the material *buddhi.* The real self, as pure awareness, cannot be experienced as an object of consciousness, since all direct objects of consciousness are held to be specific structures of *sattvic prakṛti* illuminated by *puruṣa.* Thus the constellation of subjective experience which comprises Hume's 'empirical' self is held to be a mere material pattern, which receives 'transcendental' illumination from *puruṣa.*

Indeed, the habitual identification of the individual subject with the illuminated transformations of *buddhi* is traditionally diagnosed as a particular mode of *buddhi,* and is held to be the root form of *avidyā* or ignorance regarding one's true nature. A standard illustration of the purported relation between the real self and the ego is the following. The sun, as reflected on the surface of the ocean, appears to fluctuate with the waves and agitations of the surface, but it is only the reflection, rather than the actual sun, which is affected by these disturbances. Thus the phenomenal self undergoes innumerable changes, and possesses the varied structural attributes of its physical medium, but these are true only of the reflection of self in an external substrate. According to the Sāṅkhya-Yoga account, the real self is not to be identified with the Cartesian realm of qualitative experience, and there simply is no *res cogitans,* no thinking substance in Descartes' sense. It is material substance which thinks, and it is the self which makes these thoughts conscious.

3 The Distinction Between Consciousness and Content

The foregoing brief sketch of the Sāṅkhya-Yoga view presents a striking contrast, on a number of points, with the traditional western analysis of self and mind. For my present purpose, the most salient point of contrast concerns the theoretical separation between consciousness and mental representation, rather than questions concerning the proper analysis of the self. In the ensuing discussion, I would like to examine some consequences of a general mind/consciousness distinction, as inspired by the ancient Indian *darśana,* but I would like to examine these consequences from the more contemporary perspective of mental representation in cognitive science and AI. Thus not all of the following claims and considerations are held to be contained in, nor to be strictly deducible from, the traditional Sāṅkhya-Yoga philosophy; rather I wish to explore some features of interest which follow from a broad conceptual alternative to the Cartesian picture, the latter of which still colors much contemporary thought on these subjects.

On the Cartesian view, subjective presentations or 'ideas', as *structured objects,* are held to belong to the non-material realm. But, at the same time, they are held to represent external material objects and states of affairs. This leads to an immediate and very serious difficulty, since if the two spheres are metaphysically independent, then it is highly unclear how events in the mental realm could be related to or systematically represent objects in the material realm. Thought becomes separated, by an ontological chasm, from the external objects of thought, and the inner world of subjective experience threatens to become a solipsistic bubble. If some sort of (divinely harmonious) correlation between the two spheres is advocated, in order to provide a foundation for epistemology, then this leads to a profligate and unaesthetic version of metaphysical repetition, where objects in *res extensa* are gratuitously mimicked by their representational duplicates in *res cogitans*.

Mind/consciousness dualism, though obviously not free from difficulties, at least possesses a theoretical advantage over traditional western dualism in this respect, since the conscious aspect of subjective experience is entirely disengaged from its semantical or representational form. The epistemological and cognitive advantages of this separation are most clearly illustrated in the 'base level' cases of veridical perception, and singular thoughts involving objects encountered in the environment. In these cases, representational structures are of the same metaphysical category as the objects they represent, just as a chair and a photograph of a chair are both composed of matter. This metaphysical (and hence causal) homogeneity provides a theoretical foothold for naturalistic accounts of representational content, since, just as the photograph of a chair can be accounted for *via* the assorted optical and chemical processes involved in its formation, so the representational content or structure of the perception of the chair can (in principle at least) be accounted for in terms of the physical interactions and processes involved in *its* formation.

It is the conscious, subjective aspect of visual perception which serves to motivate the introduction of a distinct metaphysical category, not the causally induced representational structure of perception, since it is theoretically feasible that the latter can be explained in terms of *unconscious* mechanisms, of generally the same sort that would be applied in the case of robotic 'vision.' The deep philosophical problem in the case of human perception lies not in the explication of structure, but rather in the fact that this bio-mechanically induced structure is imbued with conscious awareness. Thus Descartes' dualism of mind and matter, wherein conscious presentations *qua* structured objects are placed in the purely non-material realm, embodies an unhappy conflation of theoretically distinct features, and it thereby creates a causal abyss which insulates perceptual experience from the material objects which are perceived. It is consciousness, rather than content, which provides the most compelling impetus for dualism.

Thus the Sāṅkhya-Yoga division is based on a more perspicuous distillation of the salient components of subjective experience, where the world of qualia, the 'Cartesian theatre' of conscious presentation, is analysed as a mixture rather than as a realm of pure substance. The representational content of qualia is due to the respective material structures or patterns which comprise them, while conscious illumination is external, and ontologically independent of structure. Conscious illumination of structure is made possible by the particular *composition* of the pattern, wherein the transparent material of thought-stuff enables the light of *purusha* to enter. Thus representational content is comparable to the geometrical patterns of a stained glass window, which exist independently of their illumination by the sun. The illuminated patterns are a composite, consisting both of structured glass, which is analogous to the thought-shapes assumed by *sattvic buddhi*, and of luminescence, which is the consciousness of *puruṣa*. In this manner, qualia are seen as composite phenomena, and, in the case of perceptual experience, there is no ontological chasm separating representational objects from the objects represented. Instead, it is the underlying conscious illumination, common to all experience, which defines the metaphysical boundary line between subject and object.

It is important to note that the foregoing critique of the Cartesian picture is not based on a general philosophical tenet to the effect that a metaphysical homogeneity *must* obtain between a representational object and the object represented, and which would entail that the mind must be material if it is capable of representing material entities. This inversion of Berkeley's argument against matter is far too stringent, and indeed would face serious difficulties in the case of abstract and non-specific objects, universals, nonexistents, etc. Rather my point is simply that representational content, even for a cognitive phenomenon as basic as visual perception, is not susceptible to naturalistic explanation, if this content is held to reside in a non-material medium. This point is intimately linked to the earlier observation concerning mental causation with respect to bodily motions, since the problem of interaction between disparate substances cuts both ways. However, if mind and environment are held to belong to the same metaphysical realm, then mental content can both cause and be caused by other physical events. This at least opens the door to explaining mental representation and the evolution of cognitive structure through appeal to the interaction between an organism and its environment, while it is not at all clear that this door is open on a Cartesian account.

These considerations suggest that the ancient Indian separation between conscious awareness and representational content has interesting ties to current issues in cognitive science and philosophy of mind. As will be discussed below, a crucial relation between human and computer-based intelligence is

brought into sharp focus by mind/consciousness dualism, while this same relation is obscured by the familiar Cartesian divide. The research programmes of cognitive science and AI are based on a computational paradigm, in which it is assumed that cognitive phenomena, both natural and artificial, are founded on computational procedures instantiated in physical systems. The material view of mind endorsed by the Sāṅkhya-Yoga analysis is in principle quite compatible with this paradigm, since it is conceivable that the cognitive activities of *manas-buddhi* can best be described in computational terms. *Manas-buddhi* is an unconscious mechanism which manipulates the various representational structures involved in perception, cognition, and language, and from a late 20th century perspective, these manipulations can perhaps most plausibly be characterized as computational.

In this respect, there is a very strong resemblance between the activities of the cognitive organ of mind, and the syntactic manipulations carried out by a computer. Given a sufficiently sophisticated robot, *viz.* one which could perform 'autonomously' in a rich and complex behavioral environment, and thereby pass some suitably rigorous version of the Turing test, there would be many theoretical considerations in favor of attributing mentality to such an artifact. From an abstract perspective, there does not seem to be a significant difference between the mechanical activities of *manas-buddhi* and the computational procedures of an 'artificially intelligent' system. This indicates that the Sāṅkhya-Yoga position enjoys a very close fit with modern functionalist accounts of mind.

The fit is especially close in that the Sāṅkhya-Yoga analysis of mind makes no allusion to the presence of conscious experience. I think that, conversely, it is quite clear that functionalism and the computational paradigm have nothing important to say about consciousness, and that subjective experience is an element which is theoretically extraneous to the research programmes of cognitive science and AI.[5] Computational procedures are mathematical abstractions, defined over purely formal objects, without reference to consciousness, and there is little reason to believe that consciousness or genuine subjectivity will simply 'emerge' from instantiated procedures of sufficient complexity or subtlety. Indeed, it is only the representational structure of qualitative experience which is relevant to its functional role, and hence this role can always be filled by an object with the same functional structure but which by hypothesis is *not* conscious. Since

---

5 Various 'missing qualia' arguments, such as Block (1978) and Jackson (1982), have objected that functionalism is not in principle able to account for subjective experience, and I would agree with the view that considerations of functional structure alone cannot provide a sufficient condition for consciousness. On the dualistic picture exposited in the present work, functional structure is not even a sufficient condition for the ability of an instantiated system to *receive* consciousness, since receptivity is a feature specific to composition.

there is no contradiction inherent in this last claim (unless one is smuggled in by begging the question), it follows that consciousness *per se* is theoretically superfluous to functionalism.

4 Cartesianism and the Chinese Room

Searle (1980) provides a well known and important criticism of the classical AI position, a position which asserts that computational models have explanatory significance with respect to genuine mental states. Searle employs his celebrated Chinese room thought experiment in an attempt to demonstrate that instantiating an abstract system of rule-governed symbol manipulation is *not* a sufficient condition for the attribution of mentality. In brief, Searle argues that programs are merely recipes for grinding syntax, while genuine understanding requires an intrinsically semantical or intentional component. Thus no system can possess genuine understanding simply in virtue of instantiating a program, and therefore algorithmic processes cannot provide an adequate account of the mind. He insists that *human* understanding is inextricably tied to the particular physical 'stuff' out of which brains are made, and thus functional structure alone cannot account for intentionality, since functional structure is by definition independent of its medium of instantiation.

In the context of the present discussion, it is interesting to look a bit closer at Searle's claim that the physical stuff of which brains are made is crucial to their ability to sustain the intentional aspect which underlies true understanding. In (1990), Searle further clarifies his position by explicitly equating intentional content with the ability to be consciously entertained by a subject. This equation sheds a very revealing light on the Chinese room scenario, because it shows that what is most obviously missing from the Chinese room is *conscious presentation of representational content*, which Searle in turn takes to be essential for genuine understanding. I think that the intuitive force behind Searle's objection to the doctrine of strong AI lies in the fact that the Chinese room lacks conscious presentations, and specifically, it lacks presentations of the representational content associated with the system's ability to pass the Chinese Turing test. Searle, as homunculus in the room, only has presentations regarding his understanding of English, and of his attendant ability to execute the program, written in English, which tells him how to manipulate the 'meaningless' Chinese symbols. This lack of 'semantical presentation' with respect to the Chinese symbol processing constitutes the real asymmetry between English and Chinese, and thus the pivotal issue in the thought experiment is conscious presentation, which we may or may not wish to equate with 'real understanding.'

So, according to this reading of Searle's argument, lack of conscious presentation can be isolated as the crucial element missing from the Chinese

room, and failure to reproduce the specific material composition of the brain can be isolated as the *cause* of this lack. Hence potential isomorphism of functional structure has been granted, while genuine mentality (*i.e.*, presence of associated conscious presentations) has been denied, on grounds which ultimately reduce to the specific material composition of the brain. This gives rise, on the one hand, to a striking similarity between the Searlean critique of strong AI and the Sāṅkhya-Yoga analysis of mind, although it also highlights a crucial difference between the two, and reveals a strong neo-Cartesian element in Searle's position. The resemblance will be discussed first.

As described in the previous section, it is a basic tenet of the Sāṅkhya-Yoga view that the specific composition of the mind's thought-structures is essential to whether they can receive conscious illumination. It is only the very special *sattvic* characteristics of *buddhi* which allow human thoughts and experiences to 'absorb' consciousness. Thus, according to a functionalist rendition of Sāṅkhya-Yoga dualism, the actual medium of instantiation is essential to whether representational content can support conscious awareness. It is the particular compositional details of human, brain-supported thought-stuff which make awareness of content possible, and to this extent, Sāṅkhya-Yoga dualism is in accord with Searle in maintaining that isomorphism of functional or representational structure is not a sufficient condition for reproducing subjective experience, and thus that the Chinese room cannot possess all the characteristics essential to the human mind-self complex. Functionalist methodology is capable of reproducing the mechanical characteristics of mind, which, when suitably instantiated, would allow an artificial system to completely simulate all behavioral aspects of human beings. If the mind is viewed as computational, while consciousness is held to be independent of computational structure, then it is possible to maintain simultaneously that the research paradigm of cognitive science should be capable of accounting for all phenomena which are properly cognitive or mental, *and* that functionalism does not provide a sufficient condition for the presence of qualitative experience.

In contrast to the Searlean view, mind/consciousness dualism would deny that conscious presentation has any intrinsic link to representational content. Representational content is defined in purely functional terms, and those terms, once instantiated, are sufficient to account for the causal and behavioral interactions of all 'intelligent' physical systems, including human beings. In this manner, there is no properly *cognitive* difference between human qualia, and functionally isomorphic structures exploited by unconscious machines. If mind and consciousness are separated and the mind is construed computationally, then there can be no appeal to a uniquely 'semantical' component in the characterization of genuine understanding.

Computational or formal procedures are by nature syntactical, even though the structures and manipulations involved may be *interpreted* as playing representational or semantical roles. For example, biological or design based heuristics may, for various reasons, attribute representational content to the structures over which computations are performed, where these reasons will advert to patterns of actual or potential interaction between the system and its environment. This is an operational account of content which does not correspond to the sense of 'intrinsic aboutness' which Searle attempts to capture through his appeal to conscious presentation. However, on the model of mind/consciousness dualism, conscious awareness can not in itself imbue a mechanical system with 'intrinsic aboutness.' The fact that some human representational structures are consciously presented does not give them the 'magical' power to uniquely determine their (operationally specified) objects in the external world. Hence consciousness does not confer a privileged semantical dimension to the mind, but, on this account, merely illuminates some of the syntax.

So it would appear that any appeal to conscious presentation, in order to explicate the type of 'real' mental content which purportedly distinguishes the human mind from functionally equivalent artifacts, constitutes a tacit invocation of *res cogitans*. Of course, most contemporary proponents of this Cartesian heritage would not be willing to posit a distinct ontological realm in which subjective ideas reside, but rather would hold the familiar components of mental substance to be somehow exuded by relevant organic processes, thereby endorsing a 'naturalized' version of the Cartesian picture. Yet in many ways this constitutes an unfortunate move, since the severe problems which afflict Descartes' notion of mental content indicate that this notion is better abandoned rather than naturalized. For example, if genuine semantical or representational content is equated with the potential to be consciously entertained, then the deeply troublesome issue concerning privileged introspective access to the mental realm is re-introduced, and in fact assumes central theoretical importance, since potential awareness then becomes an essential criterion for meaning. This problem is especially acute in an attempted naturalistic setting, since it is not clear that the introspective awareness of meaning has *anything* to do with the causal processes responsible for physical behavior, or with the many irredeemably *unconscious* processing states of the brain which intercede between the various episodes of conscious thought.

Even on the physicalist hypothesis that consciousness is nothing but a product of material brain state, it still does not follow that these particular brain states are of unique causal or functional significance. Still less does it follow that they are causally significant *by virtue of* the physical properties responsible for consciousness. Thus there does not appear to be a naturalistic or causal motivation for the identification of 'genuine' representational con-

tent with the Cartesian realm of conscious presentation. And this leaves open the possibility that, even on a physicalist account of awareness, the semantical content which is presented has little or nothing to do with the explanatorily significant brain states which govern both the physical behavior of the system and the continued evolution of the internal sequence of brain states. Indeed, within a naturalistic setting, it is difficult to see why the presumed self-transparency of *res cogitans* should have any causal significance, even with respect to the transition from one conscious state to the next.

Mind/consciousness dualism would of course deny that consciousness has privileged introspective access to content (since content is independent of awareness) and it would also deny that consciously presented content, as such, has any unique semantical or causal status. A central feature of this model is that there is no intrinsic representational or functional difference between conscious and unconscious mental processes. Instead, there is a material continuum out of which representational structures are composed, where the composition of thought-forms can be more or less transparent, and hence more or less susceptible to illumination, just as ordinary translucent objects can retain local variations in opacity. Conscious illumination takes place only with respect to a small portion of the causally efficacious material mind, and is not an intrinsic property of mental activities. Thus consciousness does not enjoy infallible or perfectly self-transparent access to the realm of content, nor does conscious presentation itself affect the functional role of the structures illuminated.

5 Conclusion

So, in a variety of respects, the ancient Sāṅkhya-Yoga version of dualism provides a more felicitous dividing line between substances than does the Cartesian parsing of mind and matter. Descartes' picture of consciousness and the mind has had a tremendous influence on the development of western thought; to a large extent it still defines the terms in which the mind/body problem is conceived, and it colors many contemporary notions of mental content and representation. In this sense, I think that the Sāṅkhya-Yoga view is at least as worthy of serious philosophical attention. A far wider range of cognitive phenomena are made available to naturalistic explanation on the Sāṅkhya-Yoga account, simply because the mind is included in the physical world. But, at the same time, the unique and autonomous status of conscious subjectivity is preserved. Of course, dualism of any flavor tends to be theoretically distasteful, and this is equally true of the Indian picture. Thus a primary motivation for examining this ancient philosophical system lies not in defining yet another type of polarity between substances, but rather in providing a different framework within which to consider issues such as the mind/body problem, which will in turn serve to articulate the dilemmas

that a satisfactory form of monism must address. By conceiving the problem along slightly different lines, the Sāṅkhya-Yoga view already solves some of the difficulties which would haunt a naturalized version of Descartes' model.

References

Āraṇya, H. 1983. *Yoga Philosophy of Patañjali.* State University of New York Press; includes a translation of Patañjali's *Yoga Sūtra* and Vyāsa's commentary.

Dasgupta, S. 1963. *A History of Indian Philosophy.* Vol. 1, Cambridge University Press.

Block, N. 1978. 'Troubles with Functionalism.' *Perception and Cognition: Minnesota Studies in the Philosophy of Science.* Vol. IX, W. Savage (ed.), University of Minnesota Press.

Feuerstein, G. 1980. *The Philosophy of Classical Yoga.* St. Martin's Press.

Jackson, F. 1982. 'Epiphenomenal Qualia.' *Philosophical Quarterly* 32, 127–36.

McDowell, J. 1986. 'Singular Thought and the Extent of Inner Space.' Appearing in P. Pettit and J. McDowell (eds.), *Subject, Thought and Context.* Oxford University Press.

Raju, P. T. 1985. *Structural Depths of Indian Thought.* State University of New York Press.

Searle, J. 1980. 'Minds, Brains and Programs.' *Behavioral and Brain Sciences* 3, 417–24.

______ 1990. 'Consciousness, Explanatory Inversion, and Cognitive Science.' *Behavioral and Brain Sciences* 13, 585–96.

Sinha, N. 1979. *The Samkhya Philosophy.* Oriental Books Reprint Corporation, New Delhi; includes a translation of the *Sāṅkhya-Kārikā* of Iśvarakṛṣṇa.

This page intentionally left blank

Rensselaer Polytechnic Institute

The Self in Advaita Vedānta*

Eliot Deutsch

> I must first know myself, as the Delphian inscription says; and I should be absurd indeed, if while I am still in ignorance of myself I would be curious about that which is not my business. . . . Am I indeed a wonder more complicated and swollen with passion than the serpent typho, or a creature of a gentler and simpler sort. . . ?
> (*Phaedrus*, 230a, trans. Jowett)

THE GOAL of philosophy for Advaita Vedānta, like that expressed by Socrates, is self-knowledge: it is to discover whether indeed man is, in essence, merely a complicated passionate being or whether he is also a simpler being whose nature is centered in a divine reality. This quest for self-knowledge is pervasive in Indian thought and is given a preeminent place in Vedānta. For in the Upaniṣads, which are the main source of systematic Vedānta, it is held that a knowledge of the self is a "saving" knowledge; that he who knows himself knows reality and overcomes all pain, misery, ignorance, and bondage (e.g., *Muṇḍaka Upaniṣad*, III, 2, 9).

What is the self, the knowledge of which yields freedom and wisdom? How does it relate to what we ordinarily take to be our self—our physical organization, our mental activities and capacities, our emotional and volitional life? Is the self essentially individual or is it universal? What, if any, are its limits of awareness and knowledge? These are the questions which Advaita Vedānta, the non-dualistic school of Vedānta, expounded primarily by Śankara (c. 8th-9th centuries A.D.), is committed to answer.

The purpose of this paper is to explore just how these questions are answered by Advaita Vedānta, and to interpret these answers systematically. Since the developments of "depth psychology" and

* I would like to express my thanks to the American Institute of Indian Studies (W. Norman Brown, President) for giving me the opportunity to spend a year in India (1963-64) to do research in Advaita Vedānta.

"existential-phenomenological analysis," some of these answers are not as unfamilar to us as they were, comparatively speaking, a short time ago. I have, however, for the most part, avoided making comparisons with western thought, except when greater clarity can be had from such comparisons, for the reason that the Advaitic answers to these questions about the self are not, in the spirit of that philosophy, to be determined, appreciated, or tested by comparing them with other possible answers, but rather by comparing them, as it were, with human experience itself. The Advaitic analysis, in short, is "true" if it enables us to acquire a fuller realization of who or what we essentially are; it is "true" if it describes adequately, and brings coherence to, our experience of ourselves. And this truth is determined, according to Advaita Vedānta, only if we relate ourselves to the dynamics of the system itself.

I. Paramātman

> Bāṣkala asked Bāhva three times about the nature of Brahman: the latter remained silent all the time, but finally replied:—I teach you, but you understand not: silence is the Ātman.[1]

A person's essence is unapproachable through his name: and in the Spirit, in the Absolute where pure silence reigns, all names are rejected, and hence the truth of self cannot be reached through a name.

The application of a label to someone too often implies that in the deepest ontological sense that someone is a "something." Labels, according to Advaita, are mere conventions and sounds; and to disown all labels in a penetrating inward intuition means not only to recognize the "nothing" that one is, it means to become the silence, the "everything" that alone is.

Ātman, for Advaita Vedānta, is pure, undifferentiated self-shining consciousness, timeless, spaceless, unthinkable, which is not-different from Brahman and which underlies and supports the individual human person.

Ātman is pure, undifferentiated, self-shining consciousness. It is a supreme power of awareness, transcendent to ordinary sense-mental consciousness, aware only of the Oneness of being. Ātman is the

[1] Śankara, *Brahmasūtrabhāṣya*, III, 2, 17; translations mine unless otherwise indicated.

name for that state of conscious being wherein the division of subject and object, which characterizes ordinary consciousness, is overcome. Nothing can condition this transcendental state of consciousness: among those who have realized it, no doubts about it can arise. Ātman is thus void of difference, but it is not for Advaita simply a void: it is the infinite richness of spiritual being.

> This Atman is self-luminous.[2]
> There being nothing else but the Ātman, what should it see or know in particular. . . .[3]
> Wheresoever there is doubt, there the wise should know, the Self is not. For no doubts can arise about the Self, since its nature is pure immediate consciousness.[4]

Ātman is timeless. It cannot be said to have arisen in time, to be subject to a "present," or to have an end in time—for all such sayings apply only to that which is relative and conditioned. Time, according to Advaita, is a category only of the empirical or phenomenal world. Time, with its before and after, can make no claim on the "eternal Now" which is the state of Ātman realization.

> This intelligent Self is without birth or death.[5]
> Whenever the nature of the Self is referred to by scripture, it is said to be eternal, unchanging and without any origin.[6]

Likewise it is spaceless. Spatial relations hold between objects of the empirical order; they cannot be extended so as to constrain that which is the content of spiritual experience: "It is impossible for the body to be the receptacle of the Self."[7]

It is unthinkable. "The eye does not go there, nor speech, nor mind."[8] "It is great and self-effulgent, and its form is unthinkable."[9]

Thought functions only with forms, in multiplicity; Ātman, being without determinate form, cannot be an object knowable by the mind, perceivable by the senses. Thought is a process; Ātman is

[2] *Bṛhad-āraṇyaka Upaniṣad,* IV, 3, 9.

[3] Śankara, *op. cit.*, II, 3, 18; trans. by Vinayak Hari Date.

[4] Sureśvara, *Naiṣkarmya Siddhi,* III, 37; translated by A. J. Alston (London: Shanti Sadan, 1959).

[5] *Kaṭha Upaniṣad,* I, 2, 18.

[6] Śankara, *op. cit.*, II, 3, 17.

[7] Sankara, *Kaṭhopaniṣadbhāṣya,* II, 2, 12.

[8] *Kena Upaniṣad,* I, 3.

[9] *Muṇḍaka Upaniṣad,* III, 1, 7,

a state of being. Thought objectifies; Ātman is pure "subject" which underlies all subject/object distinctions. "The knowledge of the Ātman is self-revealed and is not dependent upon perception and other means of knowledge."[10]

It is not-different from Brahman. "Tat tvam asi."[11]

Identity judgments such as expressed in the *mahāvākya* (great saying) "*Tat tvam asi*"—"Thou art that"—are not, for the Advaitin, mere tautologies: they are the concrete representation of a movement of thought from one ontological level (of particularity), through another (of universality) to yet another (of unity), wherein the attainment of the latter negates the distinctions between the former. One begins with individual consciousness (*tvam*), passes on to universal consciousness (*tat*) and arrives at the pure consciousness which overcomes the separative reality of both the individual and the universal, and which constitutes their ground.

Tat tvam asi means the affirmation of a common ground, viz., consciousness, to the individual and to "God." The identity is obtained by stripping away the incompatible or contradictory elements of the "that" and "thou" and thereby arriving at their common element or basis. According to the Advaitin, if we were to take the term "thou" strictly as the finite individual person, and the word "that" as a transcendent deity, then clearly an identity between them is meaningless; by definition they are two different things. But there is a "secondary sense" of the terms which discloses the proper meaning of the proposition.[12] If the "thou" denotes the pure consciousness underlying human being, and the "that" the pure consciousness which is the ground of divine being then a complete identity between them does indeed exist. The individual self, apart from all the factors that differentiate it from pure consciousness, is the same as the divine, apart from its differentiating conditions. *Tat tvam asi*, according to Advaita Vedānta, is a significant proposition provided we deny the reality of the two terms as such and grasp their basic underlying identity.

In the depth of my being, then, I am not-different from Reality: the depth of my being, which is not "mine," is Reality. Man, according to Advaita, is not a mere conditioned being, so that if you were to strip away his desires, his mental activities, his emotions

[10] Śankara, *Brahmasūtrabhāṣya*, II, 3, 7.
[11] *Chāndogya Upaniṣad*, IV, 8.
[12] This method of stripping away contradictory elements of individuation in order to arrive at underlying unity is called *jahad-ajahad-lakṣānā*.

and ego you would find a mere nothing; he is spirit, he is consciousness, he is free and timeless being.

Ātman cannot be an object of thought, and it cannot be arrived at as the conclusion of a rational argument. In order, however, to orient the mind towards it and to prepare the mind to accept it as a fact of experience, the Advaitin does proffer rational arguments, the most common of which is a sort of *cogito ergo sum* argument. Śankara states that "to refute the Self is impossible, for he who tries to refute it is the Self."[13] No one, Śankara maintains, can doubt the existence of Ātman, for the act of doubting implies the very being of the doubter, who must thereby affirm his own existence. Vidyārana expresses the argument thus:

> No one can doubt the fact of his own existence. Were one to do so, who would the doubter be!
>
> Only a deluded man could entertain the idea that he does not exist....[14]

The argument is not without its difficulties. A subtle and unsupported transition is made between the Ātman and the *jīva* (the individual conscious being) so that the argument does not so much prove the Ātman as it does the *jīva*. It is the *jīva* which has the kind of self-consciousness described in, and presupposed by the argument, and not the Ātman, which is pure consciousness. In other words, by establishing the self on the basis of the inability to deny the doubter who would deny the self, the self that is being established, apart from any other difficulty in the argument, is necessarily a qualified one; it is the self in waking consciousness who is aware of an 'I' and which, as will be shown, is associated with a qualified reality, the *jīva*, and not with Ātman, the non-dual Reality.

The main purport of the argument, however, as suggested, is to orient the mind towards Ātman: its pragmatic success then is dependent upon one who enters into a dialogue with it Ātman.

II. Jīvātman

1. *The Phenomenal Status of the Jīva*

> The individual soul is not directly the highest Ātman, because it is seen to be different on account of the limiting adjuncts; nor is it different

[13] Śankara, *op. cit.*, II, 3, 7.

[14] *Pañcadeśi*, III, 23-24; translated by Hari Prasad Shastri (London: Shanti Sadan, 1956).

from the Ātman, because it is the Ātman who has entered as the *jīvātman* in all the bodies. We may call the *jīva* a mere reflection of the Ātman.[15]

The individual human person, the *jīva*, is a combination of "reality" and "appearance." It is "reality" insofar as Ātman is its ground; it is "appearance" insofar as it is identified as finite, conditioned, relative. The individual self then is empirically real, for it is a datum of objective and subjective experience; but it is transcendentally unreal, for the self, in essence, is identical with the Absolute.

In attempting to understand the status of the *jīva*, Advaita Vedānta proffers two theories or metaphors, both of which as indicated above were suggested by Śankara.[16] A common intent informs these two theories, but the differences between them are instructive. According to the first, which is called *pratibimba-vāda*, the theory of reflection (and which is associated primarily with the Vivaraṇa school of Advaita), the *jīva* is a reflection of Ātman on the mirror of *avidyā* ("ignorance"), and as such is not-different from Ātman in essence. Just as in everyday experience where we know that the face in the mirror is not really different from the face in front of it, that it does not have an independent life of its own, and yet we maintain a distinction between them, so the *jīva*, reflected in "ignorance" is not really different from its prototype, the Self, and yet continues to be a *jīva* until the mirror itself is removed. The *pratibimba*, the reflection, is actually as real as the *bimba*, the prototype, being in essence the same thing, and is misjudged to be different only because it, the *pratibimba*, appears to be located elsewhere than the *bimba*. One attains the truth of non-difference, then, the moment that one understands that one is a reflection of Ātman which only appears to be different from it, but which is in reality identical with it.[17] And just as the reflection of a person in a body of water varies according to the state of the water, according as the water is calm or turbulent, clean or dirty,

[15] Śankara, *op. cit.*, II, 3, 50, translated by Vinayak Hari Date.

[16] Some interpreters of Advaita (e.g., V. P. Upadhyaya) have suggested that three theories to explain the non-dual status of the *jīva* and its phenomenal appearance are to be found in Advaitic literature: the *pratibimba-vāda*, the *avacchedaka-vāda*, and the *ābhāsa-vāda* (of Sureśvara). The differences between the *pratibimba-vāda* and the *ābhāsa-vāda*, however, are subtle and were not emphasized to any extent in the literature as a whole.

[17] Cf. B. K. SenGupta, *A Critique of the Vivaraṇa School* (Calcutta: S. N. SenGupta, 1959), pp. 240-41.

so the reflection of the Absolute varies according to the state of *avidyā* (ignorance) upon which it is reflected. The minds of men vary: some are more, some are less under the influence of passion and desire; some are more, some are less capable of intellectual discrimination and insight. The Absolute appears differently according to these differences among individuals.

The first description of, or metaphor about, the appearance of the *jīva* has this advantage: it suggests that the clearer the mirror the more perfect is the relation between the *jīva* and Ātman. As the mirror loses its individual characteristics it reflects better that which is presented to it. The *pratibimba-vāda* suggests, then, that rather than being restless, anticipating, desiring, our minds ought to be as a clear and calm mirror capable of reflecting truth.

The second theory (which is associated with the Bhamati school of Advaita) is called *avacchedaka-vāda*, the theory of limitation. According to this theory, consciousness which is pure and unqualified, without sensible qualities, cannot be "reflected," and hence the analogy with the mirror, when pressed to the point to which the *pratibimba-vādin* takes it, breaks down. The individual is not so much a reflection of consciousness as he is a limitation of it; a limitation which is constituted by the *upādhi* of ignorance. The term *upādhi*, translated usually as "limiting adjunct" or "limiting condition," is frequently employed in Advaitic analysis. A *upādhi*, a limiting adjunct or condition, generally means the qualification or limitation of one thing by another thing. For Advaita, in the context with which we refer to it, a *upādhi* is the limitation, due to mental imposition, of infinity by finitude, of unity by multiplicity. It results in the seeing of the Infinite by and through limitations or conditions which do not properly belong to it. So long as ignorance exists the individual is not seen as he really is, but as he is separated from other individuals, as he is conditioned and finite. Just as space is really one but is seen through limitations as if it were divided into particular spaces, the space in a pot or a jar, so the Self is one but is seen through limitations as if it were multiple. The limitations, grounded in ignorance, are only conceptual: the Self is essentially unlimited and real.[18]

[18] The objection frequently raised against this position is that it is impossible (unintelligible) for the *jīva* to be a limited appearance of Ātman and at the same time to be the locus of *avidyā*; for it becomes then both the product and the source of the limiting adjuncts. For purposes of understanding the Advaitic position on the nature of the *jīva*'s status, however, it is not necessary to examine the subtle dialectical interplay between these two rival theories, for the "theories" are in fact but metaphors and anal-

Avacchedaka-vāda, the theory of limitation, gives a somewhat greater empirical reality to the *jīva* than does *pratibimba-vāda*, the theory of reflection, in this sense, that whereas the *jīva* in the *pratibimba-vāda* is a mere mistake, as it were, in the *avacchedaka-vāda* it is a necessary "practical" reality. Insofar as we are subject to ignorance, to *avidyā*, it is necessary, as a matter of practical convenience that we perceive individual persons and objects as separate, distinct realities. For both theories, however, the *jīva* qua *jīva* is an illusory appearance. The existential status of the individual human person, whether as a reflection of Ātman or as a limitation of Ātman, is one of qualified reality; its essential status is that of unqualified reality, of identity with the Absolute.

2. *The Nature of the Jīvātman*

The analysis of the empirical or phenomenal self in Advaita Vedānta is, as we interpret it, primarily a "phenomenology" of consciousness. It is a description of the kinds of awareness that one has of oneself when one is subject to *avidyā*. Advaita Vedānta does not so much explain the self as it describes the process by which we come to believe that it exists. In other words, the Advaitic analysis of the empirical or qualified self is concerned to answer this question: By what process of mis-identification do we form the belief in the reality of what is in fact an illusory appearance?

The first domain of our self-awareness, according to Advaita, is that of our being awake in relation to a world of external objects. We take ourselves as an 'I' or 'me' insofar as we are aware of things, events, and processes. Waking consciousness (*jāgarita-sthāna*) or primitive I-consciousness is "intentional": Vedānta agrees with existential phenomenology in this, that waking consciousness must have objects in order for it to be; that being awake means being *awake to*.[19] And this being awake is never static. Waking conscious-

ogies which seek to clarify rather than to demonstrate the ultimate non-difference between the *jīva* and Ātman and the phenomenal status of the *jīva* as *jīva;* and being metaphors they cannot be pressed to the point where they can stand as independent theoretical statements. In other words, the Advaitin of whatever persuasion begins with the central experience of non-duality and looks for ways by which this can be communicated and for what follows from it; he does not arrive at the fact of non-duality as the conclusion of an inductive or deductive demonstration.

[19] Or as Sureśvara puts it: "There is no manifestation of the 'I' without a modification of the mind directed to the external." *Naiṣkarmya Siddhi*, III, 58.

ness ceaselessly shifts its attentive energies to different objects and to different aspects of a single object; it lapses into disinterest and is freshly stimulated by other interests. Waking consciousness is thus time-bound: when in it we do not see things all at once nor do we think about things in a comprehensive totality; we are constrained to a successive representation, to a considering of things one at a time.

According to the *Māṇḍūkya Upaniṣad*, wherein the analysis of the states of consciousness of the self is given, waking consciousness is directed towards the enjoyment of gross objects.[20] It is involved in desire and consequently is dissatisfied; for no object is capable of sustaining our interest in it, no object is capable of fulfilling the demands that we make upon it. We are aware of ourselves initially, then, as a waking, time-bound and enjoyment-seeking, but dissatisfied being. This self which is so dissatisfied, according to Vedānta, is the self which has identified itself with its physical body. In waking consciousness we are aware that we are a part of physical nature, that we are formed from it, are sustained by it, and that eventually we return to it. The *Taittirīya Upaniṣad* speaks of this kind of identification as the *annamayakośa*, the sheath or vesture of food (matter).[21] In waking consciousness we think that "this physical body is me"; and we miss thereby the truth of the Self. The physical body (or what becomes termed the *sthūla-śarīra*, the gross body) cannot be the Self, for it is conditioned, temporal, finite.

The first stage in the development of self-awareness, or the starting point in the phenomenology of consciousness, is that of waking consciousness, the basic standpoint of all philosophy. Advaita is not concerned here to deal with waking consciousness functionally, to explain, that is, the manner of its functioning in relation to its

[20] "The first quarter is Vaisvanara (all-beings) whose sphere of action is the waking state, where consciousness relates to external things, and who possesses seven limbs and nineteen mouths, and who enjoys gross objects (*jāgaritasthāno bahisprajnāḥ saptaṇga ekonimśatimukhaḥ sthūlabhugvaiṣvanaraḥ prathamaḥ pādaḥ*)" I, 3.

Śankara, in his commentary on this, attaches no importance to the seven limbs (*saptanga*); he says that they are incorporated only for the purpose of completing the imagery of the Agnihotra sacrifice as set forth in *Chāndogya Upaniṣad* (V, 18, 2). The "nineteen mouths"—the five senses, the five organs of action, the five vital forces, and the fourfold division of the internal-organ (*antaḥkarana*), however, are looked upon as the avenues by which we are led to, and become enjoyers of, gross objects. They are psycho-physical constituents of the self which bring it into meaningful contact with external objects and which makes possible its attachments to them.

[21] *Taittirīya Upaniṣad*, II, 2.

contents; it is concerned only to identify it as such and to point out the kind of self-identification that follows from it. In waking consciousness the self has forgotten itself. Whereas in reality the self is Ātman, pure "subject," it here takes itself as a mere object. It is some-thing among other things; it is "estranged" from the fullness of its being.

The second level or state of consciousness distinguished by Vedānta is that of dream-consciousness (*svapra-sthāna*).[22] It is the self as inwardly cognitive of the impressions carried over from the waking state: it draws its sensuous and passional materials from past experience and in turn influences future experience within the waking state. The dream state is the state of fancy and wish-fulfillment and functionally, if not in content, it may be equated with the "subconscious" of analytic psychology. Śankara states that "after the cessation in sleep of the activities of the senses, (the individual) creates a subtle body of desires, and shapes his dreams according to the light of his own intellect."[23] The dream state is thus a natural extension of the waking state but, as further pointed out by Śankara, "the waking state lasts until such time as the knowledge of Brahman is attained, while the dream state is daily sublated by waking life." In other words, the stuff of dreams, although rooted in the waking life, is not empirically real in the same sense as the content of waking life, for one recognizes it for what it is, namely, mere fancy and desire.

Advaita Vedānta attaches a good deal of importance to the phenomenology of dream consciousness in order to show the continuity of consciousness and the persistence of self-awareness throughout all states of consciousness. The independence of the subject of experience from its object is clearly exhibited in the dream state. No matter how deeply involved one is with the objects of dream one retains an independence from them (as embodied in the judgment 'I had a dream'), and indeed a greater freedom with respect to them than is possible in waking consciousness. One can violate all the rules of spatio-temporal relationships that hold between empirical objects when one is in the dream state; and one can readily "transpose" one's emotional relations to these objects from one kind of past experience to another. But one is involved in one's dream, as well as being a witness of it, and hence one has made a fundamental mis-identification of oneself with the contents of one's quasi-liberated

[22] *Māṇḍūkya Upaniṣad*, I, 4.

[23] Śankara, *Brahmasūtrabhāṣya*, III, 2, 4; trans. V. H. Date.

consciousness. These contents are called subtle elements (*tanmatras*); not only are they less gross than the objects of waking consciousness, they subtly influence the whole fabric of waking consciousness, constituting as they do the material out of which so much of "personality" is formed. And it is precisely this subtle influence which makes it so difficult for the *jīva* to know itself, to become fully aware of the forces which motivate it and which mold its attitudes and values. The contents of the dream state are elusive, and, from the standpoint of the experiencer, they are largely involuntary: they present themselves without conscious control or selectivity. Further, in so far as the waking state is intimately bound-up with the dream state, it too is largely involuntary. A complete freedom to choose the contents of consciousness is thus denied the *jīva*. The Advaitin would agree with Spinoza, then, that we think we are free when we are aware of our desires, but that in fact we are not free insofar as we fail to understand the causes of these desires—which causes, for the advaitin, are traceable to a root cause of *avidyā*.

The three *kośas* or sheaths which are associated with the dream state of consciousness and which constitute the "subtle body" (*sukṣmasarīra*) of the self are, according to Vedānta, the *prānamayakośa*, the sheath of "vitality," the *manomayakośa*, the sheath of "mind," and the *vijñānamayakośa*, the sheath of "understanding." The *prānamayakośa* is the self identified as a vital being. We not only identify ourselves initially as gross physical beings, but also as animate beings. The *Taittirīya Upaniṣad* regards this vital aspect of our being as consisting of five breaths[24] which pervade the body and give power to the senses, and to the organs of action[25] which likewise constitute this vesture of the self. Although the physiology here is no doubt crude, representing as it does only the most basic awareness of the nature of vitality, the analysis does point to a furthering of self-awareness in the direction of the true self and to the kind of mis-self-identification that we make. It has been shown in modern psychology (namely by Freud) that all of us believe that "death is what happens to the other person," that we cannot imagine what it would mean for our own consciousness to cease, that in short we believe ourselves to be immortal. This belief is grounded

[24] Identified as the in-breath (*prāna*), the out-breath (*apāna*), the diffused breath (*vyāna*), the body space (*ākāsa* = *samāra*) which is the "context" of the vital forces, and the support (*puccha* = *udāna*) for these functions.

[25] Identified as speech, hand, foot, anus, genital.

in the *prānamayakośa*, in our identifying ourselves as persistent vital beings. But there is no genuine—that is to say, spiritual—immortality for the *jīva* qua *jīva*: it is condemned to "existence," to phenomenal reality until such time as it realizes Ātman.

The *manomayakośa* and the *vijñānamayakośa* comprise the mental life of the self. They are made up respectively of *manas* and the five sense organs, and *buddhi* and the five organs of perception. A general distinction is made throughout Indian philosophy between two aspects, functions, or domains of mental life: that of *manas* or "sense-mind" and *buddhi* or "intellect." The precise meaning and role given to these two domains of mental life differ from system to system, and in the different stages in the development of a single system, but a common enough core of meaning is present throughout. *Manas*, the sense mind, is an instrument, sometimes taken as a sense organ itself, which assimilates and synthesizes sense impressions, and thus enables the self to make contact with external objects. It is involved with these objects and consequently gives rise to the possessive 'my' and 'mine.' *Manas*, however, is somewhat blind, lacking as it does a discriminating objectivity, and is thus referred to as a mental condition of doubt (*vikalpa*). *Manas* rushes out through the senses into the form of any object presented to it and assimilates that form to itself. It furnishes the self with precepts which must be acted upon, transformed and guided by a higher mental function. This is the work of the *buddhi*, the intellect or reason. The *buddhi* is an instrument of discrimination, a faculty of judgment; it determines our intellectual attitudes, fortifies our beliefs, and makes understanding possible. Whenever one is aware of oneself, then, as a rational being who is capable of intellectual insight and judgment, one is involved in this *vijñānamayakośa*. Both the *manomayakośa* and the *vijñānamayakośa* form the *antaḥkarana* or "internal organ," which is the psychological expression for the totality of mental functions in waking-dream consciousness.

According to Vedānta, all the mental activities of the self who identifies himself as a mental being are subject to a pervasive *avidyā*. No true self-knowledge is possible until such time as Ātman is realized, until such time as there is a fundamental change in consciousness. The self, being subject, cannot be an object to itself. All knowledge which takes place apart from Ātman is within the subject/object situation: the self as knower, within the context of the waking and dream states of consciousness, is a self which necessarily separates itself from the object of knowledge and hence always imposes something of itself upon the object. In short, with *manas*

and *buddhi*, with the totality of mental functions, we may be self-possessed but never possessed by the Self.

The next stage of consciousness identified by Vedānta is that of deep sleep (*suṣupti*), which is characterized by a bliss (*ānanda*) which follows from the holding in abeyance of all distinctions.[26] Its *kośa* counterpart is the *ānandamayakośa*, the sheath of bliss, and is referred to as the *kārana-śarīra*, the causal body of the self. Deep sleep consciousness is not "transcendental consciousness," the spiritual consciousness in which oneness is attained, but it is not on that account to be construed as a void.[27] Defined initially in negative terms as an absence of objects, of desires and activities, it is then described in positive terms as a state of joyous consciousness. It is, writes Śankara, "an abundance of joy caused by the absence of the misery involved in the effort of the mind. . . ."[28] Deep sleep consciousness is the self as unified and integrated; it is not so much an overcoming of the distinctions which make for activity and desire as it is a harmonization, and a being a witness (*śaksin*) of them. Distinctions, in other words, are not abolished or transcended here, rather they are present in a kind of pure potentiality; and for this reason the *jīva* is said to perceive here pure *avidyā*, the cause or source of all distinctions. Further, deep sleep consciousness with its counterpart self-identification, the *ānandamayakośa*, is also called "causal"; it is the ground for future actions: the various sheaths that are manifest in the waking and dream states are latent here ready to unfold as prompted by the effects of one's past experience (*karma*).[29]

The *Māṇḍūkya Upaniṣad* describes lastly, through a series of negations and affirmations, that pure state of consciousness associated with Ātman.

> They consider the fourth to be that which is not conscious of the internal world, nor conscious of the external world, nor conscious of

[26] *Māṇḍūkya Upaniṣad*, I, 5.

[27] The Advaitin argues that in the state of deep sleep consciousness is present, that deep sleep is a state of consciousness and not of non-consciousness, although there are no objects there with which it relates or interacts. This is the case because upon returning to waking consciousness one does affirm "I had a wonderful sleep," and if consciousness were absent altogether in that state no memory affirmation of it would be possible. Consciousness thus persists even in the absence of all of the instruments of sense and cognitive experience.

[28] Śankara, *Bhāṣya* on *Māṇḍūkya Upaniṣad*, I, 5.

[29] Cf. Vidyāraṇa *Pañcadeśī*, XI, 69-72.

> both the worlds, nor a mass of consciousness, nor simple consciousness, nor unconsciousness; which is unseen, beyond empirical determination, beyond the grasp (of the mind), undemonstrable, unthinkable, indescribable; of the nature of consciousness alone wherein all phenomena cease, unchanging, peaceful and non-dual.[30]

This is "transcendental consciousness," the attainment of which leads to self-realization, to freedom (*mokṣa*).

In later Advaita Vedānta, a distinction (already present but not as sharply distinguished in early Advaita) is clearly brought out between two types or kinds of spiritual experience related to "transcendental consciousness"; that of *savikalpa samādhi*, the borderline experience as it were between the *jīvātman* and *paramātman*, and *nirvikalpa samādhi*, the pure spiritual experience of Reality.[31] The distinction between these two is subtle but important, as it brings out with such clarity the uncompromising non-duality of Advaita Vedānta and makes clear, in terms of the distinction between *savikalpa samādhi* and *suṣupti*, the deep sleep state, the fundamental difference between *paramātman* and *jīvātman*—from the standpoint of the *jīva*.

Savikalpa samādhi means "determinate" spiritual experience. In it an awareness of duality is absent, but unlike in *suṣupti*, the deep sleep state, the emphasis is not so much on the *absence of duality* as it is on the *presence of non-duality*. Whereas in *suṣupti* the self is still conscious of itself as a knowing subject, although there is nothing there as such to be known, in *savikalpa samādhi* the self is aware only of the presence of Reality. In *suṣupti* all the phenomenal activities of the self are suspended in a kind of serene blank; in *savikalpa samādhi* they are concentrated on the Real. In *suṣupti* the self is still very much the *jīvātman*; in *savikalpa samādhi* the self is passing into Ātman.[32]

Nirvikalpa samādhi is the consummation of that process. It is the pure "indeterminate" intuition of non-duality. In *savikalpa samādhi* the self is aware of Reality; in *nirvikalpa samādhi* it is Reality.

To sum up: the Vedāntic analysis of the self into four domains of consciousness affirms that there is no discontinuity of consciousness,

[30] *Māṇḍūkya Upaniṣad*, I, 6.

[31] We are here translating *samādhi* somewhat freely as "spiritual experience," to fit our general context of discussion. More precisely it might be called "unifying concentration."

[32] Still, "even the view that the self becomes Brahman" is only a verbal statement, for he is always Brahman." *Bṛhadāraṇyakopaniṣadbhyāṣa*, VI, 4, 6.

that there is but one consciousness, namely, that associated with Ātman, which appears in different states because of various *upādhis* or mis-identifications of self with one or more aspects of phenomenal selfhood. The four states of consciousness, then, are really stages in the development of one's power of awareness, and correspond to the ontological levels recognized by Vedānta. The waking and dream states, which can here be brought together under a single category,[33] correspond to the phenomenal world of gross and subtle bodies (termed *Virāt* and *Hiraṇyagarbha*); the state of deep sleep and of *savikalpa samādhi*, to the qualified Brahman (*saguṇa*) or the Divine (*Īśvara*); and transcendental consciousness, *nirvikalpa samādhi*, to *nirguṇa* Brahman or Reality.

And just as the ontological levels form a hierarchy of existence and value, the final term of which is not so much a part of the series as it is an annulment of the series, being incommensurable with it, so the stages in the development of consciousness form a hierarchy, the last stage of which transcends and overrides all the others. In the waking-dream state the self is caught up with objects, external and internal, and loses sight of its true nature as pure "subject." In deep sleep consciousness the self is free from objects but has not yet transcended itself. *Turīya*, the fourth, is that self-transcendence which brings about the awareness of Reality, of oneself as not-different from Reality, and the awareness of the finitude and ultimate unreality of what has preceeded it.

Phenomenally, as *jīvas*, as individual conscious beings, we are a multi-personality. We become the roles and functions that we perform; we become the kind of person we conceive ourselves to be; we become the many identifications we form of aspects of our self. Although we recognize the forces and conditions that act upon us, social as well as physical and psychological, we remain their victims: we ascribe ultimacy to them as we know of nothing to take their place. This, according to Advaita Vedānta, is the process by which we come to believe in the independent reality of the individual self, and consequently to deny the reality of the Self. Its root cause is ignorance (*avidyā*). We are ignorant in so far as we make of our physical, our bio-physical, our mental and emotional vestures some-

[33] This bringing of waking and dream consciousness into a single category is justified by the manner in which relations are established between the various stages of consciousness, by the fact that both are said to be comprised of the same "seven limbs" and "nineteen mouths," and by their ontological contexts, which together constitute the phenomenal world.

thing substantial, real, and ultimately valuable—without realizing that all being, reality, and value is grounded in, and arises from, our true Self. We are ignorant in so far as we take the expressions of consciousness, waking, dream, deep sleep, as constituting the highest development of consciousness—without realizing that our consciousness, being identical with the Absolute, knows no limits.

We have then, diagrammatically expressed, the following spectrum of aspects and states of the self:

THE SELF IN ADVAITA VEDĀNTA

	State of consciousness	*Epistemic designation*	*Ontological level*	*Sheaths of the self*	*Bodies of the self*
Param-ātman	turiya (transcen-dental) nirvikalpa samādhi	subject/object non-duality	Brahman nirguṇa (Reality)		
			Īśvara or Brahman saguna (divine being)		
Jīv-ātman	savikalpa samādhi				
	suṣupti (deep sleep)	subject/object harmony		ānandamaya (bliss)	kārana-śarīra (causal body)
	svapra (dream)	subject/object duality	Virāṭ and Hiraṇyagarbha (phenomenal world)	vijñānamaya (understanding) manomaya (mind) prānamaya (vitality)	sukṣma-śarīra (subtle body)
	jāgarita (waking)			annamaya (matter)	sthūla-śarīra (gross body)

III. Fruitfulness of the Advaitic Analysis

What has western philosophy to learn from the Advaitic analysis of the self? It can first of all, I believe, learn something from the methodology of levels as employed in Advaita Vedānta. What is true or valid to one level of our experience (e.g., to waking consciousness) is not necessarily true or valid to another level of experience (e.g., to "transcendental consciousness"). The distinction between "self" and "non-self," between "I" and "you," is a valid distinction from the phenomenal standpoint; but it is not, for the

Advaitin, a valid distinction from the standpoint of Reality. If it is the case that our "real self" is not to be identified with any aspect of our "empirical self," then these two levels of selfhood, being different in kind, are incommensurable. Realizing this, philosophy can many take a new look at arid rational-metaphysical arguments that have dogged the course of its development; arguments, for instance, about the relationship of the soul to the body (for since they are on different levels, no relation can be established by reason between them).

Second, and perhaps of greater importance, western philosophy can learn something from the spirit of quest exhibited in Advaita Vedānta. Since the Renaissance, western man has prided himself on his exploring the depths of an infinite universe without setting limits in advance to this exploration. By comparison with the Advaitic extent and spirit of exploration, however, this western claim is simply a myth. We are in fact quite convinced that our consciousness does have essential limits; and such a conviction tends to preclude any attempt to go beyond these limits. The Advaitin is a genuine "empiricist" in his approach to experience, and not the sense-chained one we associate with the term. He is willing —and the western philosopher could well follow his example—to pursue experience as far as it can take him, and to accept the implications of that experience squarely. The Advaitin is the "existentialist" who is willing to take risks, for the path which he follows is admitted to be "as sharp as the razor's edge." But if Socrates is right in believing that it is absurd to try to know anything without first knowing oneself, then taking risks is an integral part of what it means to be a philosopher.

THE CONCEPT OF THE ABSOLUTE AND ITS ALTERNATIVE FORMS

1. Philosophy starts in reflective consciousness. Reflection is the awareness of a content *as to* a mode of consciousness. The phrase 'as to' means some relation and it is in reference to this relation that the concept of the absolute has to be understood. What is this relation?

2. We ordinarily understand a relation as between the contents of consciousness and we have here apparently to conceive some such relation as subsisting between content and consciousness. Of the two relations—between content and content, and between content and consciousness—the latter has to be understood in terms of the former. Not that we know relation in objective consciousness before we come to be reflectively conscious of a relation between content and consciousness. To be aware of a relation *as such* between two terms, of their relation as distinct from their whole having the form of being, is to be aware at the same time of the possibility of the terms being not related in that relation. Now unlike a term known to be existent, a relation known to subsist between two terms cannot be *imagined* to be absent. The possible absence of a known relation, in fact, can never be apprehended in merely objective consciousness. Hence we can know a relation *as such* of two contents only in reflection which is specifically the consciousness of a relation between content and consciousness. Yet although there can be no consciousness of a relation prior to reflection, the relation of contents may be said to be prior to the relation of content and consciousness in the sense that the former is intelligible by itself while the latter is intelligible only in terms of the former.

3. There are those who hold that because we are only reflectively conscious of relation, the relation between content and consciousness should be read into the relation of content and content and not vice versa. The relation of content and consciousness is to them a transparent identity-in-difference and all relation of contents is to be understood as the same relation in an implicit form. The difficulty is that the relation of content and consciousness is not explicitly or self-evidently appreciated as an identity-in-difference by reflection itself, being, in fact, claimed to be so appreciated only in a higher consciousness called absolute or speculative consciousness. The possibility, however, of this supra-reflective consciousness may be disputed and it is

necessary in the first instance to show if and how reflection itself points to it.

4. Prima facie to reflection, the relation of content and consciousness appears as an implicative distinction. We are reflectively aware of the content as distinct from and in necessary reference to consciousness and the reference is not merely verbal but is implicitly an assertion. In the case of an asserted implication between two terms, the fact asserted is a relation (or unity), the terms of which are not the terms asserted to be in implication. If A is asserted to imply B, a factual relation is meant but A is not understood as a term of the relation. A here stands for a fact which need not have the relation but is thought as having it, the fact being not altered by being so thought. This fact should be capable of being expressed by a term that does not refer to the relation and the relation should be taken as between this term and the implied term B. For every implying term, it is necessary to find such a non-implying term, if possible; and until and unless it is found, the implication has to be taken as an *indefinite* formulation of a factual relation (or unity). To reflective consciousness there is implication between the content and the consciousness of it and the implication points to a factual relation (or unity) of which one of them at least is not really a term. One of them at least stands for an unnamed something which need not be related to the other. Some factual relation then appears to be only *indefinitely* formulated as the implication of content and consciousness in the reflective stage.

5. Now if it can be shown within reflection that this indefinite relation cannot be denied to be a distinction and cannot be denied also to be an identity, a stage of consciousness in which distinction and identity of the same terms are positively known together may be taken to be demanded by reflection which cannot understand them together. That the content is somehow distinct from consciousness is obvious to reflection but identity is not so obvious. What immediately appears is an indefinite distinction between content and consciousness: the terms are undoubtedly distinct but they are not wholly distinguishable. Have we the right to interpret this vague indistinguishability as an identity?

6. The indefinite relation of content and consciousness appears to reflection at once as distinction and as not distinction but this non-distinction cannot be positively asserted as an identity. Things may come to be distinguished that were undistinguished but what are once distinguished cannot be later taken to be non-distinct without at least one of the terms being taken to be illusory. Hence to distinguish terms

imperfectly is to be able to assert their distinction but neither to assert nor to deny their identity.

7. But it may be contended that distinction and identity stand on the same footing in respect of undeniability : like distinction identity also once apprehended cannot be denied. That we distinguish what were undistinguished does not mean that we knew their *relation* of identity which is now denied. We reply that it is possible to deny the relation of identity even when it is explicitly apprehended. There is something peculiar about the relation of identity, as expressed in the dilemma: if its terms are not distinct, they are not related at all, and if they are distinct they are not identical. We cannot get rid of the dilemma and admit identity as a relation unless we hold that while the distinction of its terms is cognitively undenied, it is positively entertained only in non-cognitive consciousness. The fact A in the context M is said to be identical with A in the context N and A has the differential characters x and y in the contexts, which are presented and yet as neither factual nor non-factual. If the characters be facts, A cannot be identical in the two contexts and if they be no facts at all, there is no distinction of terms and, therefore, no relation. They have accordingly to be taken as not cognised characters but aesthetically apprehended expressions which do not preclude the cognised identity of A and yet are themselves cognitively undenied. To assert a relation is not to *assert* a distinction between its terms but only not to deny it. We may be affectively or conatively aware of a distinction where we do not intellectually disbelieve the distinction but do not also assert it. Identity then such as is knowable is a relation of terms, the distinction of which is undenied but unknown. Now what is unknown and yet speakable as felt *may* come to be known. If the differential characters *x* and *y* come to be known, the identity of A would be denied. Identity cannot accordingly be taken to be undeniable like distinction.

8. The indefinite relation then of content and consciousness should not be interpreted as indefinite identity. There is no demand, therefore, in reflection for a stage of conscious in which content and consciousness may be explicitly seen to be identical in difference. The only necessary demand in knowledge is that what appears as indefinite in an assignable respect should be definable in that respect. If further the indefiniteness is such as *necessarily* appears to a stage of consciousness, that stage is taken to demand a higher stage where the indefinite gets defined. To reflection, the relation of content and consciousness appears necessarily as an indefinite distinction only and not as an indefinite identity. The demand is for this distinction to be

defined and for a supra-reflective consciousness where this distinction can be visualised.

9. The concept of the absolute in any form is taken to belong to a supra-reflective consciousness. The possibility and nature of this consciousness have to be understood in reflection as a necessary problem. The conception of an explicit identity-in-difference of content and consciousness is not demanded in reflection. Reflection demands only a non-implicational distinction of them to define the necessarily implicational or indefinite distinction that is presented to it. The absolute accordingly has to be problematically understood within reflection as meaning not an identity but only a completely definite distinction of content and consciousness. It may be that the supra-reflective consciousness in which the reflective implication of content and consciousness is turned into a non-implicatory distinction is itself consciousness of identity in a symbolic or metaphorical sense. But in any case identity as *a logical relation* has to be definitely denied between content and consciousness before this mystic identity can be appreciated.

10. We are concerned for the present with the conception of the absolute such as is intelligible as a problem to reflection. We have to trace in detail how the indefinite distinction of content and consciousness can be defined, how their apparent identity can be denied, how, in fact, the implicatory distinction can be resolved into a non-implicatory distinction. We shall find presently that the implicatory distinction of content and consciousness varies according as the consciousness is knowing, feeling or willing. The implication is resolvable in different ways in these three spheres and the absolute is understood in the reflective stage in terms of the mode of resolution in each case. Knowing, feeling and willing will then each have its own formulation of the absolute—viz., truth, value and reality (or freedom) respectively, as will be justified later. In the reflective stage these absolutes or formulations of the absolute will be found to be un-unifiable and to be in a sort of alternation. Whether a mystical identity of the absolutes can be reached in the supra--reflective consciousness does not concern us. Our problem is to show how reflection demands a specific absolute in each case.

11. What is specifically apprehended in non-cognitive consciousness cannot be literally formulated in terms of knowledge. Philosophy which cognitively deals with the contents of all consciousness should claim to know *not* the non-cognitive contents themselves but only *that* they are non-cognitive, that we believe in contents which we do not know. It can, therefore, speak of these believed contents

only by a sort of intellectual symbolism *as though* they were known. To reflection, the implicative distinction of content and consciousness varies according as the consciousness is knowing, feeling or willing. The variations may be represented by three modes of distinction that we recognise as between known contents. If A is distinct from B, B may be simply an other, or it may be constitutive of A, or A and B may be both constitutive of C. So the content that is distinguished in reflection from consciousness may be spoken of as unconstituted by consciousness or as constituted by consciousness or as along with consciousness constituting some kind of unity. The first mode of distinction is the relation of content and consciousness in knowing and the last two will be found to appropriately symbolise the relation in willing and feeling respectively.

12. The content of a knowing act is unconstituted by the act The *particular* act of knowing discovers and does not construct the object known, even if the object be admitted to be constructed by *some* knowing. Knowledge would appear to mean that the object known is in some sense independent of it though it may be a question whether what we ordinarily claim to know is properly said to be known and to be independent of the knowing. Where the object is coloured by the particularity of the act of knowing, it may be said to be only empirically known but it is still taken to be known so far as it is believed to be independent of it.

13. The content of willing may be said to be an act, an end, or the empirical embodiment of the end—the organisation of certain objective facts (including the means) by the end. None of these is a fact that can be said to be known in the willing, believed as independent of the willing in the sense the content of knowing is independent of knowing. Each is constituted by willing in the sense that apart from willing it is nothing at all. It cannot be said that it is a future fact that is *known* in the willing. Willing is, indeed, some form of consciousness of the future but the future here, unlike the future that is said to be known, is not a fact but a contingency, not what *will* be but what *would* be *if* it were willed, not as already determined but what is being determined by the willing and as therefore apart from the willing nothing at all. Yet the content of willing is distinct from willing, distinct as what is constituted is distinct from what constitutes it.

14. Reflective knowing is consciousness of the known object as distinct from the knowing. Reflective willing is consciousness of the act willed as being distinguished by the willing from itself. Neither is consciousness primarily of the distinction itself or the distinguishing.

Reflective feeling is primarily consciousness of the distinction (or the distinguishing) of content and consciousness. To be conscious of a content *as* felt is to be conscious of the content in unity with the feeling or of feeling unified with the content. The unity of two contents means—whatever else it means—an imperfect distinction between them ; and we have to understand the unity of content and consciousness on this analogy. The unity of which reflective feeling is conscious is the imperfect distinction between content and consciousness, taken itself as a content. To be conscious of a content as known or willed is not even to be implicitly conscious of a unity of content and consciousness. The known content appears to reflection as perfectly distinct from knowing and a unity of perfect distincts cannot be reflectively conceived to be knowable. The content of willing as constituted by the willing alone cannot be understood to constitute along with the willing any unity other than the content itself. It is only in reflective feeling that we are conscious of something, viz., value that is as much content as consciousness, that is not indeed *both* of them at once but what each of them is and is not alternately. The felt content is imperfectly distinguished from feeling but not constituted by it in the sense that apart from it, it is nothing at all. We do not, indeed, know but cannot deny that the value of an object is really in it.

15. Reflective consciousness is definite consciousness of something. The content of feeling of which we are reflectively conscious is not a definite content. What is perfectly distinct from another or constituted by another in the sense of being nothing without it is definite in itself. What, however, is imperfectly distinct from another and unconstituted by it is not definite in itself. Now an indefinite can only be referred to in connexion with a definite content. What we are definitely conscious of in reflective feeling is the imperfect distinction itself of content and consciousness, this indefinite *as such* being, in fact, their unity. The indefinite content of feeling can only be referred to as a factor of this unity.

16. The content of knowing then is perfectly distinct from knowing and is unconstituted by it. The content of willing is imperfectly distinct from willing though distinct in itself and is constituted by it. Content and consciousness make a unity in the case of feeling but not in the case of knowing and willing. There are thus three modes of distinction of content and consciousness of which we are reflectively conscious. Each of these is an implicative or indefinite distinction in some sense and the indefiniteness will be found to consist in the fact that a relation that is intelligible as between content and content is

only half intelligible when taken to be the relation between content and consciousness.

17. To begin with knowing. If the content of knowing be perfectly distinct from knowing, how can the distinction be implicative or indefinite at all? It is implicative in the sense that the content of knowing is *necessarily* understood in reflection as what is perfectly distinct from knowing. The content *implies* such a reference to knowing and apparently means nothing without it. When one content is asserted to imply another, a relation is asserted of which, as we said, the implying content is not really a *relatum* but some other content which need not have the relation and which as thought with the relation is the implying content. This irrelative content may not be actually formulated but we conceive it to be formulable. It is difficult to conceive, however, what the known content apart from the knowledge-relation can possibly mean, since it is before the mind only as known.

18. The consciousness of a content as known is not, indeed, the consciousness of it as a fact necessarily related to knowing but it is not also the positive consciousness of it as only accidentally related to knowing. Reflection, in fact, on knowing tells us nothing about whether the content known is or is not necessarily known. The known content *means* indeed what is perfectly distinct from knowing but this verbal reference to knowing need not mean a *necessary factual* relation. The perfect distinction, too, does not mean that the content stands for something that would be even if it were not known by any one. Both the idealistic and the realistic positions here appear to be over-statements. In one sense, however, the realistic position may be said to have an advantage, for although reflection does not testify that what is known need not be known, it does not deny it while the idealistic view that the content is constituted by the knowing of it appears to be plainly opposed to reflective testimony.

19. It may be, however, that if and when the realistic conception of the known content as what need not be known is realised, the idealistic notion of its being constituted by knowing will be found to have a meaning. Criticism of knowledge may show that all that is actually taken to be known in the reflective stage is in some respect constituted by the particular act of knowing and, therefore, is in that respect not properly known though its known character cannot be denied altogether. Meantime we may reject the ordinary idealistic argument that to be aware of knowing an object is to be aware of recognising it, aware therefore of the object as necessarily known. There is actually no consciousness of recognising the object in the

literal sense of remembering the past knowing of it. All that is actually meant is that to be aware of knowing an object is to be aware of knowing *truth,* knowing something—in other words—to be eternal or timeless. The object may be temporal but that it is in time is not itself a temporal fact. To know is to have a timeless truth revealed but this does not mean that it is timelessly *known* and so known *again* in the present act of knowing. There is apparently no ground to assume either a previous knowing or of a timeless, impersonal or universal knowing along with the present knowing to justify the use of the word 'recognition'.

20. The realistic position that the known content may exist unknown is not denied but cannot also be asserted in the reflective consciousness. It may be said, however, that it cannot be denied or asserted because only the proposition is meaningless. Can reflection understand it as a possibility? The possibility, we reply, is understood in reference to what we take to be *self-evidently* known. Certain known contents appear at least to be self-evident, not simply evident. So long as the evident is not contrasted with the self-evident, we say about it 'it is' and not 'it is *to me*'. When it is contrasted however, we say 'it is *to me*' (though not 'to me only'), while of the self-evident we say simply 'it is'. To say 'it is *to me*' is to suggest that it may not be; and to say—in denial of this possibility—that 'it is' is to imply that it might not have been to the knower at all, that it is eternally true, that it is truth literally revealing itself. If anything then is known to be self-evident, it is known as what need not be known.

21. That the known content may exist—or more accurately—may be true without being known is then intelligible as a problem to reflection. It is, therefore, deniable, and it is not denied (though not asserted) because some content claimed to be self-evident appears to be implied in all knowledge. What is taken to be known is thus implicitly believed to be self-evident and only in this sense to be what need not be known. The realistic view of the independence of the known content, of its knownness being accidental, can only be admitted in the sense of the content being in some respect self-evident or self-revealing. The realist's definition of knowledge has to be accepted; but the question against him would be if what is claimed to be known is really known, is independent of the particular act of knowing or knowing generally. It is not the question if what is taken to be true is true. Value, for example, may be claimed to be known but it may be legitimately asked if it is known at all, if it is not merely believed in a non-cognitive way. So one may ask if what is taken by

Kant to be only empirically known as involving construction of experiences is, as the realist will claim, really known, known as independent of the constructivity of the knowing act. The Kantian view may or may not be accepted but the question is certainly legitimate if what is claimed to be known is known in the realist's sense of the term. The realist apparently would not admit the necessity of testing whether what is supposed in the first instance to be known is really known.

22. The known content should be what need not be known but where is the content that is known as such? The self-evident, we take it, is what is known to be independent of knowing in the sense of being eternally true without requiring to be known; but is anything admitted to be self-evident in the reflective consciousness? No known content is absolutely indubitable and in any case there is no agreement as to what content is indubitable. But some content or other appears self-evident in the sense of there being an explicit consciousness of doubt about it being unintelligible. The unintelligibility of doubt about a content should be distinguished from what is called the inconceivability of the opposite. Inconceivability of the opposite is understood only by trying to conceive it and to try to conceive it is to entertain a problem in thinking or meaning, if not of knowing. But there are cases where the problem of thinking the opposite of the content known is not entertained and one is conscious of its not arising at all. To be conscious of a content as self-evident is to be conscious not of its negation being unmeanable but of the problem of meaning the negation not even arising. We are conscious of the self-evident in this sense though the negation of what appears as such may later come to be conceived through a new self-evident cognition or revelation.

23. What we reflectively speak of as true or false is a judgment. The self-evident is that of which we cannot be conscious except as true. Is there then any self-evident judgment? To form a judgment is apparently always to be conscious of the *problem* of meaning its negation. This applies even to what is taken as a necessary judgment. Its negation is found to be inconceivable only after it has been tried to be conceived, and it is so found because it is but the elaboration of a systemic concept which is really a *postulate* that is neither true nor false and may admit of rival postulates. It is a direction of imagination in which what tends to be imagined is eo ipso believed. The belief is here not properly cognitive but is what may be roughly called a feeling of cognition. It is the consciousness that something *must be* without the consciousness that it is: it is as though one

dimly felt what must be. The so-called necessary judgment is an analysis of the content of such felt cognition and is not properly knowledge.

24. What however is inferred, i.e., believed as what *must be* because something *is*, is believed also as what *is* and as such cannot be said to be merely felt as known. It is only the necessary that is not consciously inferred that should be so characterised. That a conclusion follows from a premise or premises is the only content that can be claimed in reflective consciousness as self-evidently ksown and not merely necessarily known i.e., felt to be known. Properly speaking, indeed, that a conclusion follows from a premise is not a judgment though expressed as such. The 'following' is not a relation of contents that is itself a content co-ordinate with them. Still as it is not an arbitrary subjective relating of contents, we cannot say it is not a *believed* relation though it is not content of consciousness.

25. A disguised form of this inferential 'following' we have in a judgment like 'A implies B' which may be paraphrased 'to know A is to know B'. What does the word 'is' here stand for? The two knowings stand for subjective acts but the connecting word 'is' does not mean another act co-ordinate with them. Nor does it mean a relation of the contents A and B that is a third content. It appears to be a relation not of the contents but of the cognitions in respect of their content, being itself however no subjective act of cognition. It cannot be said not to be known though it is not content of the knowing of which A and B are contents. Whether it is the content of the reflective consciousness is not known in the reflection itself and so *for the reflective consciousness* it means neither the consciousness nor the content of consciousness.

26. The self-evident is to reflection a relation that *is* without being the content of a knowing that is known. 'To know A is to know B' cannot be said to be no judgment and yet here we are not conscious of the problem of meaning its negation. It may well be that the assertion is a mistake but when it is made, it is made on the basis of an immediate unquestioning belief which is yet on the reflective level. The problem of meaning the negation of a perceived content does not arise during the perception because perception is not on the reflective level. But it, is only reflectively that we say 'to know A is to know B' and yet the knowledge here is immediate so that if later it turns out to be false, it is taken to be an illusion and not merely a thinking error. Since the knowledge is reflectively immediate, there is not only no conscious problem of meaning its

negation, there is the consciousness of such a problem not arising. It is only about what is known as the *content* of knowing that such a problem arises. So it is about judgment generally (which is on the reflective level) except what we take to be the primordial judgment viz., 'to know A is to know B'. It is primordial judgment because it is the basis of necessary judgments that are the bases of all other judgments. A necessary judgment is still known mediately through the baffled attempt to conceive the opposite. This also is based on a judgment like 'to know the inconceivability of the negation of a judgment is to know the judgment to be axiomatic'. A judgment of the form 'to know A is to know B' is self-evident judgment which is implied in all other judgments. That the terms of a judgment *have* a relation i.e., appear to be related to the relation only means that to know the terms is to know the relation—which is just a judgment of this form.

27. That the copula of this primordial self-evident judgment is not known as a content requires further explanation. The judgment is, indeed, knowledge and knowledge of the contents A and B as known, but so far as it is knowledge of the relation implied by the word 'is', it is not knowledge of it as a content that is known. In the case of a judgment like 'A is B', we understand the copula as a relation that is at least partially a content on a level with A and B. But in the judgment 'to know A is to know B' the 'is' is not on a level with the cognitions of A and B, far less with A and B. It is not known as known like the cognitions and their contents in the reflective consciousness. How then do we say that it is known? Because the sentence 'to know A is to know B' is significant and what it signifies cannot be disbelieved. The import of the sentence or of the word 'is' in it is, in fact, understood only when already spoken and not in the speaking of it. When even understood as already spoken, it is not retrospectively taken as the content of that speaking and *distinct from it.* It is understood, in fact, as *spoken* but not *as spoken of.* The word 'is' in the judgment is not understood as known *content.*

28. What is understood (without disbelief) only as already spoken and not in the speaking of it is understood as known but not as known content. By 'known content' is meant what is *knowable* as distinct from the knowing of it. It may not sometimes be so known in the knowing of it but it may be known later. The import of the word 'is' in the judgment 'to know A is to know B' is not so knowable at all and may accordingly be (loosely) called a known no-content. It may be suggested that the import of the word 'is' here,

therefore, is the knowing act itself (or I) that is known by first appearing as content and then getting denied. It does not, however, appear at all as content to be negated : it is only symbolically spoken of later as content in the full consciousness that it is only verbally distinguished from the problematic knowing of it. Meantime the knowing of A and B that appears to reflection is definitely distinguished from it. What then the judgment 'to know A is to know B' signifies is known but does not appear as content to be accepted or rejected. It appears neither as known knowing nor as the content of such knowing.

29. This judgment is the only self-evident entity of which reflection is aware and the copula in it is not only not known as content but is explicitly known as what known content—including the apparent reflective content 'known knowing'—is not. The known position from which all content is distinguished—as represented by the copula here —is implied in all reflective knowledge, in all judgment like 'A is B' where the 'is' means this over and above a relation that is a content.

30. To sum up. In reflection on a content as known, the content implies knowing and is indefinitely distinct from it in the sense that it *means* this reference to knowing (though it means perfect distinction from knowing), that it does not mean independence of this reference and is not known as what is accidentally known. Reflection demands that it should be so known. The content is known to be only accidentally known i.e., to be only accidentally a content, when it is self-evident. The only self-evident of which we are reflectively aware is a judgment of the form 'to know A is to know B' or what the word 'is' in it stands for, this being meant in all judgments or knowledge on the reflective level. This then is the element in all that is reflectively known that is known as what need not be known, the element that is free from the implicational relation of the content to the knowing of it, the element that reflection demands to be isolated but cannot itself isolate. This is the absolute for knowing that demands to be freed from its immanence in the implicational distinction of content and knowing.

31. So far about knowing. We may now rapidly indicate how the implicative relation of content and consciousness is demanded to be resolved in the case of willing and feeling. What implication of content and consciousness does reflective willing present ? The willed content, as has been pointed out, is constituted by willing in the sense that apart from willing it is nothing at all. Yet the constituted content as definite in itself appears as a limitation to the constitutive willing :

willing appears to be necessarily the willing of what is foreign to itself. In the reflective stage the willed content is appreciated as real through the willing but yet as its limitation, not as its self-limitation. It is not apparent at this stage that the will limits itself to realise itself, freely commits itself to a being to annul it and become freer. Reflection demands that the limit that is necessarily constituted by willing should be realised as self-limitation in this sense. That is how the implication of content and consciousness is problematically understood to be resolvable in the sphere of willing.

32. To elaborate the problem. The relation of two contents willed at once may be taken to suggest the relation between willed content and willing. Two contents e.g., two acts are said to be willed at once when one of them is a means to the other. Two unrelated acts cannot be said to be willed at once and two acts that are jointly means to a third act cannot be *distinctly* willed in the willing of the third act. A conscious will-relation of two contents must be a relation of prior and posterior, the prior being that through which the other is possible. In the relation of means and end as *willed,* the end is the prior through the willing of which it is possible to will a content *as means.* Now this relation of means and end has to be understood as the relation of willed content to willing. A willed act has to be taken as a means to the willing of it though it is through the willing that the willed act is possible. It is difficult to understand, however, how the willing of an act is the end of the act willed, how in other words willing is realised by the putting forth of the act, how one can be said to act in a particular way in order to act freely.

33. That we objectively act to be subjectively free, that the good will and nothing but the good will is the value for which we will an act—the view, in fact, of Kant—may be called the idealistic view in this connexion. The realistic view here then would be that we act for an objective end and not for the subjective end of being free; and an extreme form of the view may be conceived that we objectively act in order that we may objectively act for evermore. In the case of knowing, we pointed out, the realistic view is prima facie more acceptable than the idealistic view. In the case of willing apparently, the idealistic view is acceptable in the first instance. The realistic view amounts to saying that there is no willed act that is good in itself and the view that we objectively act in order that we may objectively act is just its logical consequence and its reductio ad absurdum. The idealistic view is consonant with the nature of reflective willing—viz., that the prior is the end, but in the reflective stage it appears to

amount to the barren statement that it is good to will what it is good to will. The demand is to find a vital meaning for the statement, to understand how willing is a willing of itself.

34. Reflection indicates how it is to be understood. There is the difference between 'I will this act' and 'I will this act that I ought to will', when the former is not contrasted with the latter, the will-consciousness is expressible as an imperative 'let this act be done'. When, however, it is contrasted, it is expressed as mere information 'I am doing it' while the latter is properly expressed as an imperative 'let this be done' which implies 'I may not do it'. In contrast with it, the former appears as a statement of fact though by itself it appears as an imperative. The implication is that willing as represented by the former is itself a being that is to be superseded by willing as represented by the latter, that what is taken as free is implicitly not free, that therefore we will an act in order to get rid of the being of the act, get rid of the self-complacent will to continue in this being—what may be called the will to indolence. Reflection indicates the way but cannot understand willing or freedom at its limit.

35. In reflective feeling there is the definite consciousness of an indefinite distinction of the content felt and the feeling of it. The content felt is not definite in itself like the known or willed content and is understood in reference to this indefinite distinction definitely appearing to reflection *as though* it were a unity. We say 'as though', because the unity does not appear as a definite self-subsistent unity from which its constituents are distinguishable. The implication of content and consciousness would be resolved in the sphere of feeling, if a unity of this kind could be apprehended. In the feeling of two contents together, we can reflectively apprehend a self-subsistent unity. Reflection accordingly demands such a unity of felt content and feeling but cannot itself understand it.

36. To explain. In the apprehension of an object having spatial parts as beautiful, both the whole and the parts are felt but differently. The whole as a known content is distinct in itself and even if it be conceivably distinguished *from* the parts, the parts *as in the whole* cannot be distinguished from the whole and, therefore, can only be imperfectly distinguished from one another. The parts are, however, distinguished in the feeling way from the whole, being, in fact, felt as not felt in the way the whole is felt. The whole is felt to be beautiful but the parts are felt to be indifferent, felt to be 'only known', such feeling of their mere knownness being necessary for the appreciation of the beauty of the whole. To feel two contents at once then is to

feel their unity and to feel them otherwise than the unity. We try to understand the relation of the felt content and the feeling of it on this analogy. They appear imperfectly distinct and unity means to reflection an imperfect distinction of the constituents. But the unity does not appear to reflection self-subsistent in the sense of the constituents being distinguishable from it. Such a self-subsistent unity is, however, demanded.

37. As in the case of knowing and willing, so here one may imagine an alternation of realistic and idealistic or objectivistic and subjectivistic views. The unity of felt content and feeling may be understood as content that is indefinitely other than consciousness or as consciousness that is indefinitely other than the content. Value may be regarded as a kind of object though not as completely distinct from the consciousness of it as the known object or it may be taken to be an impersonalised feeling as somehow expressed in the object and thus objectified symbolically but not as a known character. Neither view appears to have any advantage over the other. Does reflection, however, indicate how consciously the alternation may be stopped and the unity come to be definite in itself ?

38. Taking value, the unity of felt content and feeling, realistically as objective, we are reflectively aware of the value as referred to the known object that is distinguishable from it. We feel the object as known to be not the value, to be neutral, felt otherwise than the value. This is *feeling* the distinction of the object from the value of it and it is thus that the value tends to appear distinct in itself by having the object distinguished from it. So also if we take the value as an impersonalised feeling rather than as an objective character, in understanding it as expressed in the object we may feel the knowing of the object as somehow *inside* the impersonal feeling, involved in it without being confused with it, much as the image in a mirror shows the mirror to be unaffected by it. Thus we may be said to be aware of the felt object as well as of the individual feeling as distinct from their unity.

39. So far the relation of felt content and feeling appears to be similar to the relation of two contents of the same feeling. In both there seems to be a unity from which its constituents are felt to be distinct. There is, however, discrepancy in an important respect. In appreciating a beautiful object where we feel the object as a whole as well as the parts, the parts that appear neutral are *together* felt to be distinct from the whole that appears beautiful. The beautiful object as beautiful thus appears in feeling as a whole isolated not only from

other objects but also from its own knowable parts and thus shines as a self-subsistent something in the air. We cannot understand such a thing about the unity of content and consciousness called value. The felt content and the feeling consciousness are only *alternately* distinguished from the value that is their unity. The demand is for them *together* to be feelingly distinguished from value so that it may shine in isolation as a self-subsistent unity. To put it more correctly, the demand is for the concept of a value that is independent of valuation, of its reference to a known object. Reflection understands the problem but does not see the solution.

40. In all these cases then, reflection tries to understand the relation of the content of a mode of consciousness to the consciousness on the analogy of the relation of the contents of the same consciousness and finds that the analogy can be definitely extended only halfway, although it can indicate the way in which the indefinite aspect that remains over could conceivably be defined. In the case of knowing, how the known content could be without being known is not reflectively understood although the consciousness of the self-evident as distinct from the evident is consciousness of such a *possibility*. In the case of willing, it is reflectively indefinite how willing is the end of the act willed but the consciousness of the *ought* is consciousness of the possibility of getting rid of the being of the act willed. In the case of feeling, value should be understood but cannot be reflectively understood as the definite self-subsistent unity of felt content and feeling, but there is the consciousness of its possibility in the reflective appreciation of a beautiful object as what 'never was on sea or earth', isolated from its knowable relations not only to other objects but also to its own parts.

41. In all these cases, again, it will be noticed that what demands to be understood cannot actually be understood because such understanding would involve a species of negation that to reflection is unmeaning. In the case of knowing a known content has to be understood as what need not be content of any knowing. In the case of willing, the reality of willing has to be understood as the negation of being, of the being even of the content that is willed. In the case of feeling, the being of value has to be understood by the known as such being distinguished from it. To reflection, for the known content to be without being known is for it to be intrinsically unrelated to knowing in the way of distinction or identity. To be conscious of A and B, not to be able to deny either and yet to say they are utterly unrelated is apparently to state a contradiction. So to understand the willed

act to be distinct from willing and yet to be nothing apart from it, or—to put it differently—for the act not to *be* and yet to be real as willing is to conceive the negation of an emergent distinct or the emergent distinction of a negation. To take the felt value also as a being from which known being is distinct is to be definitely conscious of an indefinite, to entertain without disbelief an *appearance* or an indifference of being and non-being. All the three—unrelatedness, negation of the emergent or the emergence of negation, and the indifference of being and non-being—imply unmeaning modes of negation to reflection ; and yet the specifiable indefiniteness in the content of reflection demands to be defined precisely through such modes of negation. To admit the absolute in any form is to admit a negation that is unintelligible to the logic of the understanding.

42. The absolute may be generally defined from the standpoint of reflection as what is free from the implicational dualism of content and consciousness. There are three ways in which this freedom can be understood. The content may be freed from its reference to consciousness i.e., from its contenthood. Or consciousness may be freed from its reference to content, in which case it ceases to be conscious of anything beyond itself. Or the implicational relation itself may be freed from its terms as a definite self-subsistent unity. The known that is free from its contenthood is known as the content that need not be content, is the self-evident *is,* just what we call *truth*—the absolute for knowing. Consciousness that is free from its content (in the sense that it solely constitutes its content, makes the content a content, creates its distinction from itself) is *freedom* of the will—the absolute for willing. The implicational relation of content and consciousness that is freed from their distinction as a unity is *value* in itself—the absolute for feeling.

43. Truth, freedom and value then are absolutes for knowing, willing and feeling. It is impossible to avoid this triple formulation of the absolute though the notion that there are three absolutes would be just as illegitimate as the notion of there being only one absolute. The absolute is not a *known content,* about which alone the question 'one or many' has meaning. Truth is self-evident and is as such known but is no content; of value as the self-subsistent unity of content and consciousness, we cannot say either that it is not known or that it is like truth self-evident and, therefore, is no content; and there is no question even of knowing freedom, the belief in freedom being, as Kant pointed out, no intuition but willing itself. It is meaningless therefore to cognitively assert that there are three absolutes or one

absolute. The absolute has, however, to be formulated in this triple way. Each is absolute but what are here understood as *three* are only their verbal symbols, they themselves being understood together but not *as together*.

44. For freedom, an alternative name would be reality. The real is understood either as will or what is given to will. Will means free will or freedom which though not theoretically known is not disbelieved, the belief in willing being willing itself. The expression of willing is always an imperative 'let this be done', which means no known being and, if anything, means 'let this known situation here be negated, *used* as means, melted into the future'. Willing in this sense is the negation of being and is yet real, its reality consisting in the supersession of being. This comes out more explicitly in the consciousness of the imperative as the moral *ought*. *Ought* is real as the explicit rejection of the *is* or the actual that is known, explicit annulment of the known being of the presupposed 'natural willing'. Known being may be real but reality is understood in its purity as the conative rejection of known being; or more accurately, since everything has to be understood here in conative terms, as the freedom to withdraw or abstain from 'natural willing'. Natural willing, called 'natural' in reference to the *ought*, is in itself an imperative : the person who wills says to himself 'let this be done' and not 'I am doing it'. So any willing and specially the willing to abstain from willing is unknowable freedom or reality. What is given to the will is also said to be real because either it is consumed by the will as a means to itself as end or if it cannot be so consumed, it has to be taken as an opposed will. Freedom or reality may accordingly be taken as synonymous.

45. Freedom or reality then is not known or, in other words, it is meaningless to call it truth. The true is the self-evident, that of which we are conscious as known but not as known content. Freedom or willing is not known at all or if we are aware of knowing it, we are aware of its knownness as illusory. We are only aware in one grade of willing that a presupposed willing of another grade—natural willing, as we have called it—had an illusory being or appeared to be known. What is known, however, may be either the self-evident in its purity or some relation of *given* contents—meaning contents given *to the will*—with the form of self-evidence. Every judgment, as has been pointed out, involves a self-evident judgment of the form 'to know A is to know B'. The self-evident in its purity is eternal truth and not reality : but a relation of given contents is not only said to be true but cannot also be said to be not real.

46. The real then is not true but the true may be real. So speaking of truth and value, we may say that truth is not value but value is not untrue. Just as the predicates true and false do not apply at all to freedom or reality, so the predicates valuable or worthless do not apply to truth. Truth is not felt or if it is felt, it is felt as unfelt i.e., as no value. But the predicate false applies to value in so far as the falsity of a felt value is denied though its truth cannot therefore be asserted. A value like beauty is evident but not self-evident; it appears as a content to an appreciative consciousness and appears without being disbelieved as an illusion but not as what need not have reference to the consciousness. The self-evident is the true and the evident is true if it implies self-evidence and till the self-evidence becomes explicit, it cannot be said to be false.

47. Again in respect of reality and value it may be said that while the predicates real and unreal do not apply to value, reality or freedom cannot be said not to be a value. Value is a felt being and is neither given to willing nor is itself willed into existence though it may appear as the fulfilment of willing. It is in this respect similar to truth which may shine out in fulfilment of willing in the form of attention but is not brought into existence through its causality. Reality or freedom may, however, be felt and an act that is claimed to be willed is at least not morally indifferent if it is felt to have been genuinely willed. An act, in fact, that is felt to be morally indifferent is eo ipso felt to have been not free, not willing at all, the sole proof of freedom being in the retrospective moral valuation. To say that an act is free is to take it as not valueless though it may be going too far to assert that freedom is a value. Value is a felt *being* while freedom or willing is felt as the real *negation* of a known being and can at best be the felt being of a negation (of known being).

48. Thus it appears to be meaningless to speak of truth as a value, of value as real or of reality as true while we can significantly speak of value as not false, of reality as not valueless and of truth as not unreal, although we cannot positively assert value to be truth, reality to be value and truth to be reality. Each of them is absolute and they cannot be spoken of as one or many. In one direction their identity and difference are alike meaningless and in another direction their identity is intelligible though not assertable. Truth is unrelated to value, value to reality and reality to truth while value may be truth, reality value and truth reality. The absolute may be regarded in this sense as an *alternation* of truth, value and reality.

This page intentionally left blank

Acknowledgments

Matilal, B.K. "Ontological Problems in Nyāya, Buddhism and Jainism: A Comparative Analysis." *Journal of Indian Philosophy* 5 (1977): 91–105. Reprinted with the permission of Kluwer Academic Publishers.

Potter, Karl H. "*Vedāntaparibhāṣā* as Systematic Reconstruction." In S.S. Rama Pao Pappu ed., *Perspectives on Vedanta,* Essays in Honor of Professor P.T. Raju (Leiden: E.J. Brill, 1988): 91–113. Reprinted with the permission of E.J. Brill.

Chakrabarti, Kisor. "The Nyāya-Vaiśeṣika Theory of Universals." *Journal of Indian Philosophy* 3 (1975): 363–82. Reprinted with the permission of Kluwer Academic Publishers.

Siderits, Mark. "More Things in Heaven and Earth." *Journal of Indian Philosophy* 10 (1982): 186–208. Reprinted with the permission of Kluwer Academic Publishers.

Gillon, Brendon S. "Negative Facts and Knowledge of Negative Facts." In P. Bilimoria and J.N. Mohanty eds., *Relativism, Suffering and Beyond* (Delhi: Oxford University Press, 1997): 128–49. Reprinted with the permission of Oxford University Press, New Delhi.

Kapstein, Matthew. "Mereological Considerations in Vasubandhu's 'Proof of Idealism' (*Vijñaptimātratāsiddhih) Idealistic Studies* 18 (1988): 32–54. Reprinted with the permission of *Idealistic Studies.*

Matilal, Bimal Krishna. "Causality in the Nyāya-Vaiśeṣika School." *Philosophy East and West* 44 (1975): 41–48. Reprinted with the permission of the University of Hawaii Press.

Potter, Karl H. "An Ontology of Concrete Connectors." *Journal of Philosophy* 58 (1961): 57–65. Reprinted with the permission of the Journal of Philosophy, Inc., Columbia University, and the author.

Garfield, Jay L. "Dependent Arising and the Emptiness of Emptiness: Why Did Nāgārjuna Start with Causation?" *Philosophy East and West* 44 (1994): 219–50. Reprinted with the permission of the University of Hawaii Press.

Potter, Karl H. "Freedom and Determinism from an Indian Perspective." *Philosophy East and West* 17 (1967): 113–24. Reprinted with the permission of the University of Hawaii Press.

Duerlinger, James. "Reductionist and Nonreductionist Theories of Persons in Indian

Buddhist Philosophy." *Journal of Indian Philosophy* 21 (1993): 79–101. Reprinted with the permission of Kluwer Academic Publishers.

Bastow, David. "Self-Construction in Buddhism." *Ratio* 28 (1986): 97–113. Reprinted with the permission of Basil Blackwell.

Siderits, Mark. "Buddhist Reductionism." *Philosophy East and West* 47 (1997): 455-78. Reprinted with the permission of the University of Hawaii Press.

Taber, John A. "The Mīmāṃsā Theory of Self Recognition." *Philosophy East and West* 40 (1990): 35–57. Reprinted with the permission of the University of Hawaii Press.

Chakrabarti, Arindam. "I Touch What I Saw." *Philosophy and Phenomenological Research* 52 (1992): 103–16. Reprinted with the permission of the International Phenomenological Society.

Shukla, Pandit Badrinath. "*Dehātmavāda* or the Body as Soul: Exploration of a Possibility Within Nyāya Thought." *Journal of the Indian Council of Philosophical Research* 5 (1988): 1–17. Reprinted with the permission of the Indian Council of Philosophical Research.

Larson, Gerald James. "An Eccentric Ghost in the Machine: Formal and Quantitative Aspects of the Sāṃkhya-Yoga Dualism." *Philosophy East and West* 33 (1983): 219–33. Reprinted with the permission of the University of Hawaii Press.

Schweizer, Paul. "Mind/Consciousness Dualism in Sānkhya-Yoga Philosophy." *Philosophy and Phenomenological Research* 53 (1993): 845–59. Reprinted with the permission of the International Phenomenological Society.

Deutsch, Eliot. "The Self in Advaita Vedanta." *International Philosophical Quarterly* 6 (1966): 5–21. Reprinted with the permission of the Foundation for International Philosophical Exchange

Bhattacharya, K. "The Concept of the Absolute and its Alternative Forms." In *Philosophical Studies*, Vol. II (Calcutta: Progressive Publishers, 1958): 125–43.